How to lie with statistics

像"故事书"一样精彩的数据造假曝光宝典



[美] 达莱尔·哈夫◎著

廖颖林◎译

统计数字会撒谎

美国统计专家畅销50年的伟大时代经典

中国城市出版社

译者的话

在动手写这篇序的时候，我正巧看到了一则有趣的报道，标题为“生子不是福”。报道摘录如下：“芬兰图尔库大学的萨姆利·赫利在最近一期的《科学》杂志上发表文章指出，根据他与同事对 1640 年到 1870 年间 375 名妇女教堂记录的分析结果，在前工业化时代。生育儿子会显著缩短妇女寿命，而生育女儿则会对妇女的长寿起积极作用。为了保证分析结果的准确性和确定生男生女的长期效果，研究人员把考察对象限定于年过五旬的妇女。结果表明，生育子女的数目不会影响妇女寿命，但每生一个儿子，却会平均将其缩短 34 周。”如果你是一名女性，而且又尚未做母亲，读完这则报道之后，你会有怎样的想法呢？也许你的第一反应会同标题一样——生个女儿该有多好，毕竟 34 周(将近 1 年)的生命是十分珍贵的。

但是，报道的内容是否值得信赖？支撑该报道的分析方法又是否恰当呢。

在信息爆炸的年代，我们将随时随地成为信息的接收者，散布在报纸、杂志、电视、广播中的信息。良莠并存、真伪同在，我们该如何“借来一双慧眼”，看个清楚明白呢？

20 世纪 50 年代，美国的各大媒体和宣传机构越来越重视利用统计——“这个神秘的语言” ——说话，然而大量的统计数据、统计资料由于主、客观的原因被滥用，很难起到描述事实、传递信息的作用。相反，还往往对读者形成误导。达莱尔·哈夫(Darrell Huff)，一位具有深

厚统计背景的新闻记者——发现了这一现象。他在广泛调查的基础上，从报刊、杂志、书籍中，从美国统计学会一些统计学家提供的实例中，收集了大量案例，并在 1954 年写下了“How to Lie with Statistics”一书。该书一经出版，便畅销美国，成为美国 20 世纪 50 年代的畅销书之一。并受到了当时美国各种书评杂志的好评。《管理评论》认为：“哈夫先生用如此生动的、充满人情味的方式来论述统计这个干巴巴的课题，真是一讯灵丹妙药。……我们太需要这本书了，它虽然娱乐性强、浅显易读，却十分具有说服力。”《图书期刊》如此评价：“作者和制图者倾注了全力，给大家提供了一本十分轻松活泼的读物和卡通画。它们能给你带来娱乐，又能引发思考，而且还揭穿了许多统计方法的谎言。”《大西洋》评价道：“这是一本具有善意破坏性的书，读完它后，你对于‘万能统计’的信任将大大降低。”正是由于这本书融娱乐性和知识性为一体，使它成为一本具有影响力的著作。该书出版至今，多次重印并被译成多国文字。

这本经典著作自其出版至今已将近 50 年，书中所体现的思想和方法仍然广为称道。事实上，目前在美国几乎所有知名的网上书店，仍然可以购买到此书，并一直受到读者的好评。

而且，由于该书的畅销，使得“编造虚假信息”这一命题受到了人们持续普遍的关注，如果上网查询，输入关键字“How to Lie with Statistics”，我们将发现：相匹配的查询结果不是几条、几十条，而是数千条。

本书具有独具匠心的行文结构。从第 1 章到第 9 章，

作者想像自己是一个技艺高超的惯骗，正在向后起之秀面授各种行骗的技巧。协助行骗的工具有很多，包括：有偏的样本、刻意挑选的平均数、遗漏某些重要的数据、样本的误差、统计图、平面图、不匹配的资料、混淆相关关系与因果关系以及不正确地使用资料。对每一种工具的功能和使用场合，他都一一做了详细的介绍。第 10 章是一个总结性的章节，在此章中，作者摈弃了前 9 章使用的描述手法，直接正面地阐述了“对统计资料”应该“提出的五个问题”：(1)“谁说的？”(2)“如何知道的？”(3)“是否遗漏了什么？”(4)“是否偷换了概念？”以及(5)“资料是否有意义？”通过寻找这 5 个问题的答案，读者能初步判断资料是否真实可信。原著的标题为“How to Lie with Statistics”，有多种译法。可直译为《如何利用统计撒谎》，这样基本体现了本文前 9 章的行文风格，但考虑到作者的目的是识破谎言，最终将译名定为《统计数字会撒谎》。

本书具有引人入胜的行文风格。作者文笔幽默、活泼，书中的案例充实，插图生动，十分吸引人，实在很难将它与枯燥、刻板的统计书籍联系起来。在序言中，作者将伪装统计资料的人比喻成一个在昏暗的灯光下没日没夜工作的“白衣侠客”，将本书描述成“一本如何利用统计行骗的入门读物”，类似手法的描述散布于书中，十分形象。虽然书中也出现了一些统计术语和统计方法，但在作者重说明、轻证明的描述下，即使你从来没有接触过统计，也能很好地掌握其中的思想。

本书实用性强、适用面很广。虽然书中大部分案例发

生在美国，但随着我国市场经济体制的逐步确立，在日常的经济生活中我们将接触到越来越多的统计数据和资料，例如各种证券信息、投资可行性研究报告、公司财务报告等。这些资料、数据，也需要有去粗取精、去伪存真的过程。我们又该如何进行鉴别？同样的问题已经摆在我们面前。

在对各类统计数据进行利用的过程中，证券市场最具代表性。由于利益的驱动，总会有“大胆”的公司为了抬高股价或圈钱而操纵财务报表。投资者如果报据虚假的财务数据进行投资，极可能使投资付之流水。为了对众多数据准确地辨别其真伪，了解、掌握公司的操纵造假手段是必要的。资产重组、资产评估、关联交易等都能够成为利润虚增的手段，而其中每一种手段还可以包括多种形式。以关联方交易为倒，既可以销售给控股股东和非控股子公司，也可以是不同控股程度子公司之间的销售，还可以是上市，公司溢价采购控股子公司的产品厦劳务形成固定资产。美国安然公司(Enron corp.)就是一个很好的例子。在当前安然案正在调查之际，要讲清安然崩溃的玄机几乎是不可能的。安然是金融创新的巨擘，它所发明的一系列金融创新工具及衍生产品极其复杂；而同时安然也是操纵财务报表的大师，它有意识地隐瞒了大量的重要信息及高度复杂的关联交易，安然公司的财务报表多年以来在华尔街上其实无人能够真正理解，而且几乎无人敢于挑战。从这个角度来看，安然无疑是数据陷阱大师。众所周知，中国证券市场上类似的例子也层出不穷，其高明程度虽然远逊于安然，但是这种“统计陷阱”也迷惑了众多的投资者。

随着我国加入 WTO，加入到世界经济一体化的进程中，国与国之间的经济比较也变得越来越重要。在比较过程中，对比较指标的选择会产生十分微妙的差异。下面是摘自 2002 年 3 月 20 日《财经》杂志中题为“GDP 高增长的背后”一文的一些数据。中国内地、台湾，以及韩国、美国和日本 5 个地区或国家 1994 年到 2000 年营业盈余的复合增长率分别为：9.8％、9.5％、7.7％、5.2％和-1.4％，比较这些数值发现：中国内地具有最快的增长率，其余依次降低，且日本最低，为负增长。然而，如果我们把指标换成反映经济回报率的指标时却会得到另外一番景象。同样这 5 个地区或国家，营业盈余占 GDP 的比例依次为：20.6％、31.8％、30.2％、23.3％和 20.1％，营业盈余占社会固定资产净值的比例依次为：13.5％、21.1％、15.4％、15.6％和 8.0％。比较这样两组指标值，我们不难发现中国内地的经济回报率水平偏低，在这些地区或国家中大约位于倒数第二名。综合这三组数据能较好地体现中国内地近几年的经济发展状况：虽然经济增长速度较快，但是经济效率却仍处于较低水平，这篇文章的作者客观地反映了真实情况。

但如果作者有自己的打算和想法，那么这两组指标将起到完全相反的作用，第一组指标能振奋人心，而第二组和第三组指标却让人情绪低落。

在比较过程中，选择的指标是否具有可比性也是一个值得玩味的问题。同样是失业率，我们会很惊奇地发现，中国的失业率远远低于美国和欧盟一些主要的国家，几乎可以认为处于自然失业率水平上。这是否意味着我国真的

不存在失业问题呢？对此，常识就会给出一个否定的答案，因为我们能够强烈地感受到下岗所带来的影响。我国失业率低仅仅是因为采用了不同的统计口径和统计计算方法。

“五·一”、“十·一”等假日创造的无限商机已经引起旅游公司、运输部门及住宿餐饮等部门的高度重视。如何对假日经济加以正确引导和规范成为越来越多人们感兴趣的课题。为了获得全国消费者旅游习惯的第一手资料，某公司曾展开一个抽样调查，根据已有资料：旅游总收入中来自农村消费者的收入仅占 20％以下，比例较低。该公司为简化调查，剔除比例较低的农村消费者，而将调查对象定为 25 个城市的 6400 个居民。暂且不看调查结果，仅从调查方式中，我们便可肯定地得到结论：该调查结果不足以说明全国消费者的旅游习惯。虽然来自农村消费者的旅游收入只占总收入的较小比例，但由于经济水平的差异，城市消费者与农村消费者的消费偏好总是存在较大的区别。因此，用仅包含城市消费者的样本去说明既包括城市消费者又包括农村消费者的“全国”消费者旅游习惯，明显是不合理的。

上面只是经济中的几个例子，生活中我们还可以接触到各种各样的谎言，每个人都需要对各种信息进行识别，因此。掌握书中的工具是防止受骗的可行方法。

培根曾经说过：“如果一个人以种种肯定的立论开始，他必将终止于各种怀疑；但如果他愿意抱着怀疑的态度开始，那么他必将获得肯定的结论。”我想对数据资料的判断和接收也是如此。

本书翻译过程中，得到了上海财经大学统计学系刘汉

良教授的大力支持和帮助，他在百忙之中仍多次抽空帮助校对译稿。上海财经大学统计学系主任徐国祥教授对本书的翻译、出版也给予了热情关心和帮助。上海财经大学出版社黄磊先生对本书提出了许多建设性的建议，在此一并致以诚挚的谢意。

翻译是一门遗憾的艺术。由于译者文字功底和理论水平有限，原著中的精彩程度难免打了折扣，译作中也难免存在错误和不足，敬请专家、读者批评指正。

在整个翻译过程中，译者经常会为原著中精妙的思想和流畅的语言而惊喜，虽然未必能将其原汁原味地呈现给大家，但开卷有益，希望读者能与我一起分享这份喜悦！

译者 2002 年 7 月于上海财经大学

有 3 种谎言：谎言，糟糕透顶的谎言和统计资料。

——————Disraeli

对于追求效率的公民而言，统计思维总有一天会和读写能力一样必要。

———H. G. Wells

使我们陷入麻烦的通常并非我们不知道的事情，而是那些我们知道却不正确的事情。

———Artemus Ward

整数总是不完善的。

———Ssrnuel Johnson

我需要完成一个很大的课题——统计学，但却感到我的写作功底十分有限，如果不牺牲准确性和完整性，就很难使人理解。

———Sir Francis Galton

致谢

散见在本书中的关于狡辩和欺诈的小例子经过了广泛的收集，并得到了人们的许多帮助。在我对美国统计学会发出呼吁后，许多专业统计学家——请相信我，他们和任何人一样痛恨对统计资料的滥用——从他们自身收集的资料中为我提供了大量的实例。这些人，我猜想，将乐意不被提到姓名。同样，我在许多书籍中发现了有意义的事例，主要的书籍有：Martm A.Bruinbaugh 和 I. ester S.Kellogg 所著的《商业统计学》；Hadley Cantril 的《公共观点的测定》；Willard Cope Brinton 的《统计图》；Frederick E.Croxton 和 Dudley J. Cowden 的《实用商业统计学》；George Simpson 和 Fritz Kafka 的《基础统计学》以及 Helen M.Walker 的《基础统计方法》。

绪言

我岳父刚从爱荷华州到加利福尼亚州便对我说：“你们这儿治安不好。”在他所阅读的报纸中的确是这样描述的。但是，这种报纸以不轻易忽略当地发生的任何犯罪行为，并且以对于爱荷华州发生的谋杀案用比爱荷华州本地的主要报纸更多的笔墨进行大肆渲染而闻名。

我岳父得出的这个结论是一个建立在明显有偏样本基础之上的随意的统计结论。类似于其他某些更为精致的统计结论，它也存在着证据与结果不匹配的问题。它假定报纸专栏中对犯罪行为的报道是测量犯罪率的工具。

几年前，十来个调查人员分别独立地发表了关于抗组织胺药物的数据。所有的数据都证明，在经过抗组织胺药物治疗后相当大比例的感冒能够治愈。这引起了传媒的大肆宣传，至少在广告中也是如此，并兴起了药物生产的热潮，这是因为人们长期以来对药物具有强烈的需求。但奇怪的是，人们拒绝越过统计去注意一下早就了解的事实。正如一位幽默的非医学权威人士，亨利·G.菲尔森(Henry G.Felsen)所指出的那样，服用上述药物的确能在一个星期内治愈感冒，但人们却忽略了另一个事实——即使不服用上述药物，一般而言，感冒也能在七天内痊愈。

你所见所闻的许多东西也和上面的事例一样。平均数、作用关系、趋势和图表与看上去的并不总是一致。虽然经验告诉我们“眼见为实”，但眼睛告诉我们的也许隐瞒了或者夸大了一部分事实。

统计这种神秘的语言，在一个用事实说话的社会里是如此的吸引人，但有时它却被利用并成为耸人听闻、恶意夸大或简化事实、迷惑他人的工具。在报告社会经济趋势、商业状况、民意测验和普查的大量数据时，统计方法或者统计术语是必不可少的。但如果作者不能正确理解并恰当地使用这些统计语言，而读者又并不能真正懂得这些术语的含义，那么，统计结果只能是一堆废话。

在通常的科学写作中充斥着被滥用的统计数据，以至于人们很难想到这些统计数据的背后有一个“白衣侠客”在昏暗的灯光下没日没夜地工作，就像一小撮药粉、一小瓶颜料，统计把一些重要的事实弄得似是而非。巧妙伪装的统计资料胜过希特勒的弥天大谎，虽然它也会引起人们的误解，却不会因此受到指责。

本书是一本如何利用统计以达到瞒天过海目的的入门读物。看上去，它很像骗子的行骗指南。但或许我可以模仿一个已退休的窃贼——在如何神不知、鬼不觉地撬开一把锁的方面，他的回忆录达到了研究生课程的水平——替这本书说句公道话：毕竟，骗子早就掌握了行骗的技巧，而诚实的人出于自卫也应该学会它。

第一章、内在有偏的样本

通过一个极端的例子可以马上看到如何形成有偏的样本。假设你向同胞发放问卷，问卷中包含这样一个问题：“你乐意回答调查问卷吗？”整理所有的答案，你很有可能得到下面的结论：“一个选自总体、典型的横截面”中，压倒多数的人选择了“乐意”。为了具有说服力，你还可以详细列出这个比例，直至最后一位小数。事实上，大多数持否定意见的人，已经随手将你的问卷丢进最近的纸篓中，从样本中自动除名了。哪怕最初的样本中，10 个里面有 9 个会当这种“投手”，在宣布你的结果时，你仍然会遵从惯例，忽略他们。

现实生活中，样本会按照上述方式变得有偏吗？相信我，一定会。

不久前，报纸和新闻杂志上报道：近 10 年来美国大约有 400 万名天主教徒变成了基督教徒。资料来源于由丹尼尔·A.波林(Daniel A.Poling)牧师主导的调查，丹尼尔·A.波林是教派组织《基督教先驱报》(Christian Herald)的编辑。《时代》(Time)杂志描述了整个故事的梗概：

“通过对全美基督教牧师的横截面展开调查，《先驱报》得到了调查结果。调查共发出 25000 份问卷，其中 2219 名牧师反馈回了问卷，回收的问卷显示：在过去 10 年里共有 51361 名原罗马天主教徒变成了基督教徒。根据样本推算，波林得到了全国范围的估计：近 10 年来全美共有

4144366 名天主教徒改变信仰，变成了基督教徒。主教威尔·奥斯勒（Will Oursler）写道：‘即便考虑到误差，全美范围内这一数据也不可能少于 200 万或者 300 万，而且很有可能接近 500 万。’”

虽然《时代》没能指出真相的关键之处，但是它却使我们了解到被调查的牧师中超过 90％的人没有回答，这已经值得我们向它鞠躬表示敬意了。为了彻底破坏调查结果的可信度，我们只需要指出：这个“50 万”是不可靠的，因为调查中有高达 90％的牧师没有发表看法，或许他们中大多数都早已将调查问卷投进了纸篓。

根据以上判断，我们利用管辖范围内所有牧师人数，即 181000 人——该数据就是波林博士计算时所采用的数据——进行自己的推算。由于从 181000 名牧师中抽取了 25000 人接受调查，得到了改变信仰的教徒为 51361 人，如果调查全部牧师，转变信仰的总人数应该约为 370000 人。

我们这种粗糙的方法产生了一个十分可疑的数据，但是它至少与那个在全国范围发布的数据一样“可靠”，而后者却是前者的 11 倍，是的，300 万看上去更加令人欢欣鼓舞些。

至于奥斯勒先生充满自信的那句话“考虑到误差”，好吧，如果他发现了一种方法足以弥补未知量所带来的误差，那么整个统计界都将会为之感激涕零。

在上述背景下，让我们来研究一则以前的新闻报道：“1924 级的耶鲁毕业生平均年收入有 25111 美元。”《时代》(Time)杂志在评论纽约《太阳》(Sun)报的某篇报道时曾

经指出。

哇，他们干得真不赖！

可是，等一等，这个令人印象深刻的数据到底意味着什么？是否像表面看到的那样，足以证明如果你把你的孩子送进耶鲁大学，那么在老年时你就不用辛苦地上班，甚至他年老时也不用上班？

在充满怀疑的惊鸿一瞥后，我们发现这个数据的两点可疑之处：它惊人地精确；它大得令人难以置信。

任何一群分布很广的人其平均收入能精确到以元为单位都不太可能。就算是自己去年的收入，除非全部来自薪水，否则也很难知道得如此准确。但对于年收人 25000 美元的阶层而言，多种投资渠道使得收入不可能完全来自于薪水。

而且毫无疑问，这个可爱的平均数出自耶鲁人之口。即使 1924 年他们在纽哈芬接受过良好的教育，也很难保证四分之一个世纪后，他们还能坚持说真话。当问及收入时，有些人出于虚荣或天生乐观而夸大数据；有些人却故意缩小数字，特别当涉及计征所得税问题时，往往会犹豫不决，生怕与其他文件填报的数据不符，谁知道税务员又看到了什么？也许这两种趋势——夸大与缩小将相互抵消，但这种可能性极小。一般而言。一种趋势总会强于另一种趋势，但我们无从猜测哪种趋势较强。

常识告诉我们单凭某一数据很难反映实情，这是我们得到的结论。那些实际收入也许只有 25111 美元一半的人们最终会形成如此高的平均收入，最大的误差来源在哪儿？

接下来，让我们来揭开这神秘的面纱。

这是一个抽样过程。在你所遇到各种各样的课题中，大部分统计问题的核心便是抽样。抽样的原理本身很简单。但实践中对其进行的加工导致了许多副产品，有些是不正确的。举个例子，如果你有一桶豆子，有红色、有白色。那么，红色的豆子到底有多少呢？解决的办法只有一种：数豆子。然而，用一种更简单的方法，你也可以得到红豆数目的近似结果：抓一把豆子，计算其中红豆的比例，这把豆子中红豆的比例与一桶的比例基本相同。只要样本足够大，并且具有代表性，在多数情况下，样本的信息可以很好地代表总体，但如果条件不满足，抽样的结果比一个臆想的结果好不到哪儿去——除了会形成一种十分科学精确的虚假印象以外，这种方法根本不值得推荐。然而，令人悲伤的事实是，在我们看到的或自以为知道的事物中，充斥着许多从类似样本，即有偏的或过小的或两者都满足的样本得出的结论。

耶鲁毕业生的收入也来自于样本。出于理智我们可以肯定地认为，没有人能够掌握所有仍活着的 1924 级学生的情况，25 年后，许多人已经地址不详。

并且在能够取得联系的那些人中许多人根本不会回答问卷，特别是一个内容涉及隐私的问卷。一般的问卷，5％～10％的回收率已经是相当可观了。也许这个调查的回收率会高些，但不可能达到 100％。

因此，这个收入数据是建立在一个由能够取得联系并愿意回答问卷的耶鲁学生组成的样本之上的。那么，这个样本具有代表性吗？也就是说，能否假设这个样本与样本

未被包括的人具有同样的收入水平？

那些在耶鲁大学毕业生通讯录上被注明“地址不详”的迷路小羔羊是谁呢？他们是高收入阶层吗？他们是华尔街的金融家、公司领导层，抑或是制造企业或公用事业的执行总裁？不，富人的地址是不难找到的。这个班级最富有的人，即使忽略了与校友办公室联系，他们的地址也可以通过查《美国名人录》(Who’s Who in America)或其他参考资料找到。因此，我们可以较合理地猜测，那些被遗漏的人是获取耶鲁文学学士学位之后的 25 年来没能实现自己光辉梦想的人，他们是小职员、技工、流浪者、失业的酒鬼、仅仅得以糊口的作家或艺术家……六七个甚至更多这样的人将其收人相加才可能达到 25111 美元，他们不经常在班级的联谊会上注册，仅仅因为他们支付不起路费。

又是谁会将调查问卷丢进最近的废纸篓？我们不太肯定，但是猜想这种人中大部分都没有赚到足以自夸的数目是合理的。这有些类似于第一次领取薪水的小职员，当他发现薪金条上粘着一张建议对同事保密工资的数额并不要以此作为谈资的小纸条时的心态，“别担心，”他对老板说，“我与你一样，对这么低的工资感到羞愧。”

很明显，样本遗漏了对平均收入起降低作用的两种人。现在我们可以了解 25111 美元的庐山真面目了，如果它是一个真实的数据，它也仅仅代表了 1924 级耶鲁学生中可以联系到的，并愿意站出来说出所赚数目的这个特殊群体。当然，它的真实性还建立在这样一个假定基础之上——这些绅士说的都是真话。

是否能过于轻率地做出这样的假定呢？来自抽样理论的一个分支，即市场研究的经验告诉我们，人们会说真话的假定往往是不可靠的。以前曾经搞过一项旨在了解杂志阅读量的上门调查，其中的一个主要问题是：你和你的家人阅读什么杂志？当将调查结果制表并分析后发现：大部分的人喜欢《琴师》(Harper’s)，而没有多少人喜欢《真实故事》(True Story)。

但出版商提供的数据却很明显地表明：《真实故事》的发行量是几百万份，而《琴师》只有几十万份。正如这项调查的设计者所疑惑的，或许他们问错了对象？但这不可能，因为上门调查走访了美国各式各样的居民区。惟一合理的解释是许多被调查者没有说实话，导致调查结果偏离了事实。

最后你将发现，当你想了解人们到底在读什么时，询问是无济于事的。直接上门收购旧杂志看看他们能提供什么或许能得到更多的信息。你所要做的就是点一点《耶鲁评论》(Yale Renews)和《爱情罗曼史》(Love Romances)各自的份数。即便是这种方法也并不确定，它只能说明人们曾经有过什么，而不是现在有什么。

同样，当你下次在阅读时看到普通美国人(这个词最近频繁出现，但大多数情况下是不现实的)每天刷牙 1.02 次——这个数据是我瞎编的，但它与其他任何数据一样有用——请自问一个问题：任何一个人怎么可能发现这个事实？一个妇女在看了无数宣传不刷牙者是对社会冒犯的广告之后还会向陌生人承认自己不经常刷牙吗？这个统计数据只能对那些关心人们如何谈论刷牙的人才有价值，

却根本不能反映牙齿接触牙刷的频率。

我们知道，一条河永远不可能高于它的源头。但如果在河的某处藏有水电站，却可以做到。同理，对样本研究后得到的结论不会好于样本本身。当数据经过层层统计处理，最后简化成一个小数形式的平均数时，结论似乎被确定的光环所笼罩，但只要再仔细留心整个抽样过程，这个光环就会破灭。

癌症的早期发现能否挽救生命？也许吧。但通常用于证明这一点的数据却更适合支撑相反的结论。这要追溯到 1935 年，根据康涅狄格肿瘤研究所挂号处(The Connecticut Tumor Registry)的记录，从 1935 年到 1941 年。手术后 5 年的存活率大量上升。但实际上这些记录是从 1941 年才开始登记的，在此之前的数据则是通过跟踪的形式得到的。许多病人离开了康涅狄格州，其生死与否无从得知。正如医务记者雷纳得·恩格尔(Leonard Engel)所说。所存在的内在偏差已足以“解释存活率上升的真相”。

一个以抽样为基础的报告如果要有价值，就必须使用具有代表性的样本，这种样本排除了各种误差。这就是耶鲁毕业生的收入数据失真的原因，也是许多你在报纸或杂志上读到的报道毫无意义的原因。

一位心理医生曾经报道：实际上所有的人都是神经质的。暂且不去管这种提法是否破坏了“神经质”一词的含义，我们来看看这个医生的样本。也就是说。他观察了哪些人？结果证实，他是在对他的病人进行研究后得到了这个发人深省的结论，这和代表全体人的样本可差得太远、太远了。如果一个人心理健全，他永远不可能接受心理医

生治疗。

对你所读到的东西多思考一下。你将避免接受许多似是而非的结论。

记住下面这点是有益的：无形的误差与有形的误差一样容易破坏样本的可信度。也就是说，即使你找不到任何破坏性的误差来源，但只要有产生误差的可能性，你就有必要对结果保留一定的怀疑。如果你还有一丝的疑惑，想想 1948 年和 1952 年的总统大选，它们已足够证明这一点。(1948 年美国选举时，所有民调都显示民主党的杜威会获胜，结果却是杜威败给了共和党人杜鲁门，这是美国历史上至今最大的“选举惊奇”。1952 年美国大选，新闻传媒普遍看好民主党竞选人史蒂文森，最终共和党候选人艾森豪威尔以绝对优势赢得这场选举。)

更远的例子可以追溯到 1936 年《文学文摘》(Literary Digest) 的惨败。曾经准确顶测了 1932 年美国大选的 1000 万个《文学文摘》的订阅者对 1936 年的大选进行了预测，通过电话，他们向这个倒霉的杂志编辑信誓旦旦地保证，兰登 (London) 将在竞选中脱颖而出，并且与罗斯福 (Roosevelt) 所得的票数比是 370：161。这样一个久经考验的调查群体怎么可能产生误差呢？但的确有误差。正如后来许多大学论文和报社评论员发现的，1936 年就有能力购买电话和订阅杂志的人并不能真正代表选民。至少在经济上，他们是极特殊的，是有偏的，后来证实他们中许多人是共和党的选民。该样本选择了兰登，而选民心里却想着罗斯福。

最基本的样本是随机样本，它是指完全遵循随机的原

则从总体中选出样本。总体即形成样本的母体。从索引卡片档案中将每隔 10 个的名字抽出来，从许多纸张中任意抽出 50 张；在马克特街上每遇见的第 20 个人作为访问的对象。（但需要注意的是，在最后一个例子中，总体并不是全世界的人，也不是全美国人或者全旧金山人，而只是当时在马克特街上的人。一个进行民意调查的访问员说。她选择在火车站进行调查的原因是“那儿能遇到所有类型的人”。但应该向她指出的是，有些人比如婴儿母亲的代表性不足。）随机样本的检验标准是：总体中的每个名字或事物是否具有相同的几率被选进样本？

纯随机抽样是惟一一种能有足够把握利用统计理论进行检验的抽样方法。但它同样存在着缺陷。在许多情况下，获得这种样本难度很大，并且十分昂贵，以至于单纯的经济考虑就会剔除这种方法。一个更经济的替代品是分层抽样，它在市场研究和民意调查等领域中得到了广泛的应用。

为了获得分层抽样下的样本，你需要将总体按照事先已知的优势比例划分成不同的组。这时你就可能遇上麻烦：关于分组比例的信息可能并不正确。你对访问员进行指导，以确保他们调查到一定比例的黑人，按照这样或那样的比例调查属于不同收入阶层的人，一定数量的农民，等等。而且，每一组人都要确保 40 岁以下和 40 岁以上的人数相同。

这听上去很不错，但实际上呢？在黑人还是白人的问题上，大部分时候访问员能准确判断。但在收入分组时，他会出很多错。至于农民，你如何划分一个在城镇上班又

有部分时间种地的人？甚至连岁数的问题也会引起差错，为了确保准确性，访问员会挑选那些看上去明显小于 40 岁或大于 40 岁的人进行调查。在这种情况下，由于缺少 40 岁左右的人而导致样本有偏。你不可能获胜！

除此之外，如何在各层内部获得随机样本呢？最有效的办法是准备好每一层中所有单位的名单，并以随机抽中的名单构成样本。当然，这耗资不菲。于是又转为街头调查，但由于遗漏了呆在家中的人而变得有偏；白天挨家挨户上门调查，又遗漏了上班族；转而改为晚上访问，但又不能包括那些看电影和去夜总会的人。

民意调查最终将演变为一场与误差的遭遇战。所有信誉良好的调查公司将不可避免地投入到这场战斗中。调查报告的读者应谨记这点：这场战斗永远不会取得胜利。在看到“67％的美国人反对”或其他类似的字眼时，应保留这样一个问题：67％的哪些美国人？

阿尔弗雷德·C. 金西(Alfred C. Kinsey)博士的《妇女卷》(Female Volume)也和任何建立在抽样基础之上的读物一样，书中的人虽然已经证实了是不折不扣的新潮人物，但是由于抽样过程与随机抽样实在差得太远，调查结果仍然受到了质疑。抽样名单包含了太多接受过大学教育的人（女性中该比例达到 75％）和服刑人员，这已经够糟糕的了，但是更严重的缺陷是样本极有可能严重偏向于有自我宣传性行为倾向的人，而且人们很难察觉到这个缺陷。当谈论的主题与性有关时，那些沉默寡言的、对着满怀希望的访员说不，并将自己从样本中淘汰出局的人，他们与主动站出来说出全部的家伙在性行为上存在很大的差异。

布鲁克林学院（Brooklyn College）的 A·H.马斯洛夫（A. H. Maslow）所作的一项研究表明：上述推断不仅仅是猜测。他的研究样本中包括了许多女学生，后来她们都主动参加了金西博士的访谈。马斯洛夫发现：一般而言，这些女孩子在关于性的问题上观念更开放而且有更多的经历。

当我们阅读金西的书，或者阅读任意一个近期关于性行为的研究成果时，关键的问题在于如何理解它，才能避免学习到一些根本就不是那么回事的东西。在任意一个基于抽样的研究中，这个问题都十分严重，当你将你的“大部头”或者主要的研究报告采用通俗的手法进行概括时，这个问题会更加严重 。

这里至少包括了 3 次抽样：金西博士从全部妇女中抽出的样本(第一次抽样)远不能称为随机样本，不具有显著的代表性。(但与该领域以前所做的任何工作相比，这个样本的容量较大，他的数据虽然不一定是最好的，但该数据比较重要并存一定启发意义。)同样重要的是任何一个调查问卷都只不过是所有相关问题的样本(第二次抽样)。而女士们给的答案也只是她们对于这个问题看法和态度的样本(第三次抽样)。

由谁组成调查人员也会对调查结果产生微妙的影响。第二次世界大战期间，国家民意调查中心(The National Opinion Research Center)派出两组调查人员对一个南方城市的 500 名黑人进行提问，一组调查人员由白人组成，另一组是黑人。问题一共有 3 个。

其中一个问题是：“如果日本占领美国，你认为黑人

的境况会得到改善还是变得更糟？”黑人调查组中，9％的被调查者回答“变好”，而白人调查组该比例只有 2％。回答“变坏”的比例也不相同，黑人调查组是 25％，而白人调查组则是 45％。

第二个问题是用“纳粹分子”替代“日本”，两组的结果大体相同。

第三个问题试图探寻被调查者对前两个问题的真正态度。“你认为目前致力于打败轴心国比在本国内进一步推进民主更重要吗？”黑人调查组中，选择“打败轴心国”的比例是 39％，而白人调查组则是 62％。

这是由莫名因素造成的误差，它至少告诉我们，人们在接受调查时有迎合对方说好话的明显倾向。当战争时期回答一个暗含是否忠诚的问题时，一个南方黑人对白人说了一些听起来不错但并不代表他真实意愿的话不是很正常吗？当然，区别的起因也可能在于不同的调查人员选择了不同的调查对象进行交谈。

在上述例子中，任何结果都是如此明显有偏从而导致其失去价值。你可以试着自己分析还有多少民意调查的结论，虽然并无有效的检验方法来揭露它们，但却同样有偏，同样无价值。

一般而言，民意调查都带有一定方向的偏差，就像《文学文摘》一例的偏差一样，如果你对此表示怀疑，你还可以找到许多合适的例子来证明。在《文学文摘》一例中偏差在于偏向了与一般人相比，具有收入高、受过良好教育、信息来源广、灵敏度高、举止优雅、行为保守、更多固定习惯等特点的群体。

为什么会这样呢？下面的例子将有助于理解这一点。假设你是一个被分配到街道某个角落进行调查的人员，有两个看上去符合调查要求——大于 40 岁，农民——的人向你走来，一个衣着干净整齐而另一个显得肮脏、粗暴。毫无疑问，你会向后者走去，而遍布城市其他角落的同事也会进行同样的抉择。

在自由主义者或是左翼集团中可以强烈地感受到反对民意调查的情绪。他们认为这些调查都是人为操纵的。因为结果往往不能与观念和想法相对开放的人取得一致。他们以 1936 年总统选举的民意调查为例，指出调查的结果选择了共和党人，但不久投票者就做了相反的选择。

但实际上，正如我们前面所看到的，民意调查并不一定被操纵了。也就是说，并不一定要为了制造假相而恶意扭曲结果。样本有偏的趋势可以自动地操纵结果，使其变得扭曲。

第二章、精心挑选的平均数

我相信你不是一个势利小人，而我也并不从事房地产生意。但请让我们作这样的假定，并且假设，现在你正离我所住的加利福尼亚山谷不远处的一条街上寻找想买的房子。

为了抓住这笔生意，我巧舌如簧，费尽心思地让你相信附近居民的年收入大约有 15000 美元。也许这增强了你居住于此的兴趣，不管怎样，买卖最终成交。那美妙的数字也深深刻在你的脑海。而且，既然我们已经达成协议——你有那么一点势利，在与朋友聊天时，你总会看似不经意地告诉他们现在你住哪儿。

一年左右我们又见而了。作为某纳税者委员会的成员，我正在四处奔走，为降低税率，降低财产估价，或降低公共交通费用而呼吁。我的理由是：我们支付不起各种上涨的费用，毕竟，附近居民的平均年收人只有 3500 美元。也许你会参与到我们的工作中来——这说明你不仅势利，而且还挺吝啬。但是，当听到那可怜的 3500 美元时，你也禁不住大为吃惊。到底是我现在撒谎了呢？还是一年前撒了谎？

其实两次你都不能怪罪于我，这便是利用统计撒谎的奥妙所在。两个数字都是合法的平均数，有着合法的计算方法。两个数字都代表着相同的数据、相同的居民、相同的收入。所有都是相同的，但显然至少有一个数据令人误解，足以与弥天大谎相媲美。

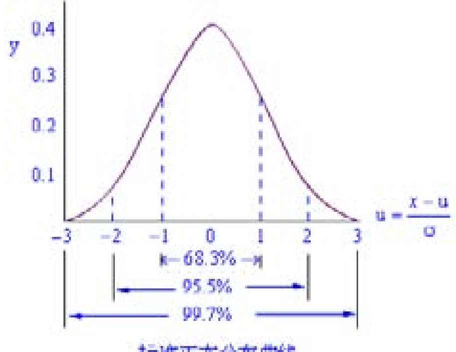
我的诡计是每次使用了不同的平均数。“平均数”这个词有很广泛的涵义。当一个家伙希望影响公共观念时，或者是向其他人推销广告版面时，平均数便是一个经常被使用的诡计，有时出于无心，但更多的时候是明知故犯。当被告知某个数是平均数时，除非你能很明确地说出它具体的种类——均值、中位数还是众数，否则对它你仍知之甚少。

当希望数据较大时，我使用的是算术平均数，即 15000 美元。将所有家庭的收入加起来再除以家庭总数便能得到。小一些的是中位数，它能告诉你一半家庭的收人超过 3500 美元，而另一半家庭的收入低于 3500 美元。我还可以利用众数，它是在所有家庭收入序列中出现次数最多的那个收入。例如，如果这个居民区中，年收入 5000 美元的家庭数远大于其他收人的家庭数，则众数就是 5000 美元。

一个没有加以限定的“平均数”，就像这个例子里的收入数据，最终是无意义的。一个可能加深混乱的因素是，有些资料的平均数落在十分接近的范围内，出于一般的目的没有必要区分它们。

当看到某些人平均身高为 5 英尺时，你便能对这些人的外形有大概的了解，而根本不需过问这个平均数到底是均值、中位数还是众数，因为它们没有过多的区别。（当然，如果你从事服装生意，并且需要为这些人赶制制服，那你就需要更多的信息，包括全距和离差，这些我将在下一章进行介绍。）当你在处理诸如上述人类特征的数据时，不同的平均数将大体相等，因为这些数据的分布十分接近正态分布。如果用曲线来描绘正态分布，你将获得一条类

似于一口钟的曲线，它的均值、中位数以及众数都落在同一点上。



标准正态分布曲线

虽然在描述人体身高时，一种平均数与其他任何平均数一样具有相同的作用，但在描述他们的经济收入时，却不是那么回事了。将一个城市所有家庭的年收入列出来，你将发现这些数据会从极小的值变动到大概 50000 美元，甚至还能发现一些极大的值。多于 95％的收入低于 10000 美元，聚集在分布的左侧。在这种情况下，分布将不再对称，也不会像一口钟，而是有偏的。它的形状类似于孩子玩的滑梯，梯子急剧升到顶点。而滑道一侧则是缓慢向下倾斜，均值与中位数离得很远。这样一来，将去年的“平均数”（均值）与今年的“平均数”（中位数）进行比较的有效性就可想而知了。

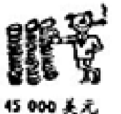
在你所买房子的居民区里，这两个平均数相差甚远，

因为收入是显著偏斜的。你的大多数邻居是小农、在附近山谷工作的工薪阶层或是靠养老金为生的退休老人，但有 3 家邻居却是来度周末的百万富翁，就是这三家的收入提高了总收入，相应地提高了算术平均数。这样一来，均值被提升到一个绝大多数家庭根本达不到的水平。这听起来像一个玩笑或者文学修辞，但实际上，几乎所有家庭收入都低于这个均值。

这就是当你听到公司执行总裁或企业所有者宣告，在他的企业中所有员工的平均收入是多少时，你应该好好思考一下的原因。如果这个数是中位数，你可以获得有用的信息：一半员工赚得比它多，一半比它少。但如果是均值（请相信我，如果没有确切指出它的种类时，多半是均值），它仅仅是企业主的 45000 美元的收入与职员低水平收入的平均数，则根本没有什么意义。“平均收入为 5700 美元”既隐瞒了 2000 美元的低收入，又隐瞒了所有者以巨额薪金的形式抽取的利润。

让我们进行更深入的研究。下图显示了有多少人获得了不同的收入。老板也许愿意用具有欺骗性的均值来进行描述：“平均收入为 5700 美元”。然而，众数更能说明问题：获得 2000 美元的职工人数最多。同样，中位数能对公司情况作进一步的阐述：一半职工收入大于 3000 美元，另一半少于 3000 美元。

这类似于双人拉锯，现实情况越糟，公司的声明却看起来越好。让我们试着用个简单的例子来说明。



45 000美元



15 000美元



10 000美元



5 700美元

5 000美元

3 700美元

← 中位数 [位于中间的数，12个数大于它，12个数小于它。]

3 000美元

← 众数 [出现最频繁的数]

2 000美元

假设你是某个制造企业的 3 个合伙人之一。这是一个收入颇丰的一年，到了年底你给本企业的 90 个职工——他们制造并且运输椅子(或者任何你所经营的东西)——共发了 198000 美元的工资，给自己与合伙人各发 11000 美元。最后，还有 45000 美元的利润可供 3 人平分。

你将如何说明这种情况呢？为了便于理解，你打算采用平均数的形式。既然所有的职工从事相同的工作，获得同样的收入，用均值还是中位数没有区别。说明如下：

职工的平均工资……………………………2200 美元

所有者平均工资及利润………………26000 美元

看上去太不公平了，不是吗，来试试另一种形式：

从利润中拿出 30000 美元以奖金的形式平分给 3 位合伙人，再将包括了所有者和职工的所有工资进行平均，的确又是平均数，结果却变成：

所有人员的平均工资或薪水………2806.45 美元

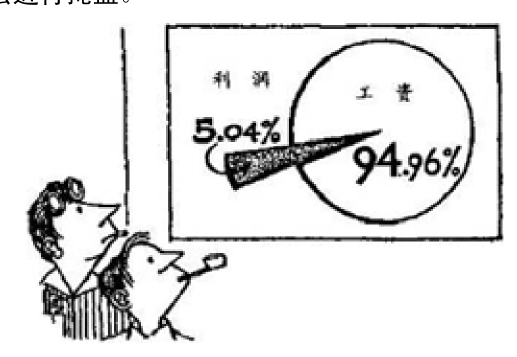
所有者平均利润……………………………5000 美元

哈，看上去不错了吧。虽然还能进一步改善，但这已经有了长足的进步，全部资金中只有少于 6％的部分形成了利润。你还可以继续如法炮制，但不管怎样，现在的结果已经足以作为公布的内容，张贴在公告栏中，或者作为与职工谈判的依据。

因为简化，这个例子是十分粗糙的，与以会计名义所做的手脚相比，它简直是小儿科。

对于一个等级森严、职员的范围从打字员到领取八十

万美元薪金的总裁的复杂公司而言，所有情况都可用类似的方法进行掩盖。



因此，当你看到一个平均收入时，首先问问：是什么的平均？包括了什么？美国钢铁公司(The United States Steel Corporation)曾经指出：从 1940 年到 1948 年间，该公司职工的平均周收入攀升了 107％。确实如此，但是，当你注意到 1940 年的数据包括更多的兼职职工时，奇妙的增长比率会大打折扣。如果你某年只工作了半年，而第二年全年工作，你的收入毫无疑问会翻番，但这与你工资率的变动无关。

你也许曾在报纸上看到，1949 年美国家庭的平均收入是 3100 美元。别太在意这个数字，除非你知道这个数字包括了哪些家庭，以及使用了哪种平均数。(甚至谁测算了数据，他如何获得这个信息以及这个数据的正确性。)

该数据正好来自普查局(The Bureau of the Census)。

如果有普查局的那份报告，那么你就可以不费吹灰之力地弄清楚其余的信息。这是个中位数，“家庭”指两个或更多具有亲属关系的人住在一起所构成的“家庭”。(如果独居也包括在内的话，中位数也许将降为 2700 美元，这又不同。)如果再同过头来读表中的数据，你还会发现这个数据建立在一个抽样调查基础之上，该调查以 19／20 的概率保证真实的数据会落在 3107 美元加减 59 美元的范围之内。对 3107 美元取整后得到 3100 美元。

类似的概率和误差范围构成了一个很好的估计。普查工作者一般都具有足够的统计知识、技术以及调查费用以确保抽样的精确度。他们并非居心叵测之徒。但并不是所有能见到的数据都产生于这样良好的环境，也并不是所有的数据会附有类似的精确度说明。在下一章，我们还将进行详细的描述。

同样，你会对《时代》杂志“编者的话”栏目中的某些项目表示怀疑。关于该杂志的新订阅者，有如下描述：“他们的平均年龄(中位数)为 34 岁，家庭平均年收人为 7270 美元。”

早期对“旧时代”读者的调查发现，“平均年龄(中位数)为 41 岁……平均收入为 9535 美元……”很自然的问题是：为什么两次谈到年龄时都指出是中位数，而关于收入却不明确平均数的类型？也许这里使用的是均值，以便利用高收入读者群来达到吸引广告商的目的。

你同样可以对第一章中 1924 级耶鲁学生的平均收人提这样一个问题：这里用的是哪种平均数？

第三章、没有披露的数据

某位统计专家曾经建议，在被告知某个调查的结果时，你需要做的就是反问一句：“为了得出这个结论，你调查了多少名被访者？”

正如以前曾指出的那样，采用严重有偏的样本几乎能够产生任何人需要的任何结果。只要样本容量足够小，或者你尝试足够多的次数，正确的随机样本也可以达到上述效果。

用户反映使用多克斯(Doakes)牌牙膏将使蛀牙减少23％！大字标题历历在目。你希望减少 23％的痛苦，于是接着往下读。这些结论出自一家信誉良好的“独立”实验室，并且还经过了注册会计师的证明，有了这些，你还想知道什么呢？

然而，如果你不是特别容易上当受骗且不盲目乐观，经验将告诉你：一种牙膏难得比其他牙膏好。那么，多克斯公司是怎样得到了如上的结论？如果他们说谎，并用大字标题报道这些谎言，他们又如何能够逃避责任呢？但事实是，根本无需说谎，下面便是更简单而有效的方法。

这里的主要把戏是不充分样本——统计角度的不充分，但对于多克斯公司而言已经足够充分了。被测试的用户，这些你只有在读小字的内容时才会发现，仅由 12 人组成。单凭这点，你便不得不佩服多克斯公司，因为它给你提供了一个冒险的机会。有的广告商索性将类似的文字略去，使得读者——即便他是一个老练的统计专家——一

头雾水：这里面到底玩了什么把戏？从这个角度来说，多克斯公司由 12 个人组成的样本还不算太坏。几年前，一个叫做可尼斯博士(Dr.Cornish)的牙粉上市了，并宣传“在治疗臼齿方面获得了极大的成功”，因为该牙粉中含有尿素，而经过实验室的证明，尿素对于治疗臼齿有极大功效。然而，值得一提的是，实验室的结论完全先入为主而且仅仅建立在 6 个案例之上。

下面，再让我们回头看看，多克斯公司可以怎样轻易地获得一个不存在漏洞并经得起检验的标题。让一组人连续 6 个月计算蛀牙数，接着使用多克斯的产品。之后只会发生三种结果：蛀牙明显增多，蛀牙明显减少，蛀牙数量无显著变化。如果发生了第一和第三种结果，那么多克斯公司就可记录下这些数字(并很好地藏于某处)，然后重新实验。由于机遇的作用，迟早有一组测试者的数据将证明蛀牙明显减少，并且这个结果足以好到作为标题并引发一场广告战。只不过不论实验者使用的是多克斯牙膏还是做面包的苏打粉，或者继续使用原来的品牌，上述结果都会产生。

使用小样本的重要性在于：在大样本的使用中，任何由于机遇产生的差异都是微不足道的，不足以作为广告标题。例如，蛀牙仅减少 2％将不会对销量有多大的提升作用。

在足够少的样本容最下，一个没有作用的结果是如何利用机遇产生的呢？——不需多少成本，你自己也可来试试。比如抛一枚硬币，多少次是头像朝上的呢？当然是一半的次数。这谁都知道。

好，让我们检验一下……我刚刚抛了 10 次，得到 8 次头像，实践证明头像以 80％的概率朝上。关于牙膏的数据也是一样的。现在，你自己试一下。也许你的结果是 5 上 5 下，但也可能是其他结果，你我的结论以同样的可能性偏离 50 对 50 的比例。但如果你足够耐心，抛上 100 次，基本上（却并不一定）能得到与半数头像朝上相接近的结果，这才代表着真实的概率。仅仅在进行了大量的实验后，平均数定律才是一种有用的描述并可用来预测。



多少才算足够呢？这又是个棘手的问题。它取决于抽样时你所研究的总体，其容量有多大，以及变动范围有多大。值得一提的是，有时样本中单位的数量看上去已足够多，但实际却并非如此。

一个典型的案例是几年前小儿麻痹症疫苗的实验。在一个社区里，450 名儿童接种了疫苗，而 680 名儿童作为

对照组没有接受疫苗。看上去，这是一个极大规模的医学实验。不久该区域感染了流行病，接种疫苗的儿童无一人患上小儿麻痹症。

对照组的儿童也无一人患上该疾病。在设计实验时，实验人员忽略了或者并没有真正理解该病的低发生率。在一般情况下，这种规模的小组中仅产生 2 名患者。因此，实验从一开始便注定是毫无意义的。15～20 倍的样本容量也许才能产生足以解释某些事物的结果。

许多医学发现同样也是急急上马，“要快，”外科大夫说，“在还来得及之前尝试用新的治疗方法。”

这不能总是归咎于医生，有时公众压力和舆论宣传也会导致一个没有经过证实的治疗方法上马，特别当需求量很大而统计背景还很模糊时。这也是伤寒菌苗几年后重新流行，而抗组织胺药物越来越多的原因。许多不成功的“治疗方法”的流行是由于疾病不确定的特性和逻辑的问题。其实，只要有足够的时间，感冒便会自行痊愈。

那么，如何避免被不科学的结论愚弄呢？是否每个人都必须成为统计专家并亲自检验原始数据？情况并非那么糟。这里有一个易于理解的显著性检验方法。简单地说，它是一种反映测验数据以多大的可能性代表实际结论、而不代表那些由于机遇产生的其他结论的方法。这便是那些没有透露的数据——假设你是个外行，不懂得其中的奥秘。但如果掌握了这个方法，你能看清其中的企图。

如果某条信息提供了显著性程度，你将对它有更深的了解。显著程度通常用概率表示，就像普查局以 19／20 的概率保证他们的结果是正确的。在通常情况下，5％的

显著水平是最低要求，有时需要更精确的 1％的显著水平，这就意味着以 99％的概率保证该结果是真实的，任何类似的事情“实践上是几乎确定”的。

还有另一类没有透露的数据，它的遗漏也同样具有破坏性。这是表明事物整体范围的全距和与平均数偏离水平的数据。在通常情况下，平均数——不管是否指明均值或中位数，都由于过于简单而导致无用。对实际情况一无所知通常比获取错误信息好得多，但有时知之甚少也十分危险。

现在美国的许多房产都是为了满足统计上的平均家庭，即 3.6 人的家庭，用现实的语言说是 3 或 4 个人，即两个卧室的房屋。这种规模的家庭，虽然是“平均”的，实际上却只代表了一小部分家庭。“我们为平均家庭建造一般规格的房屋。”制造商在这么说的同时，却忽略了具有更多人口或更少人口的大部分家庭。这样的后果是，一些地区大量重复建设两个卧室的房子，而低估了其他规模的需求。这是由不完全信息的统计资料造成巨大浪费的实例。对此，美国公共健康联合会(The American Pubic Health Association)指出：“当越过算术平均数去注意它所不能代表的实际家庭人口范围时，我们发现 3 人或 4 人的家庭仅占总体的 45％，而 35％是 1 人或 2 人，剩下的 20％则多于 4 人。”

在看到这么有说服力且权威的 3.6 人时，人们往往忽略了常识。它在一定程度上战胜了人们观察得到的事实，即许多家庭规模比之小，还有相当一部分比之大。

几乎以同样的方法，《杰塞的准则》(Gesell’s Norms)

中遗漏的数据给许多父母带来了痛苦。假设一对父母在《星期天》(Sunday)图画副刊等地方读到“孩子”将在某月份学会坐直的内容时，他们立刻会联想到自己的孩子。如果恰恰他们的孩子在该月份不能坐直，父母一定会得出结论：自己的孩子智力低下、不太正常或这很不公平等。既然一半的孩子在那时都坐不直，那就会有一半的家长将为此苦恼。当然，从数学的角度而言，这些不快将与另一半聪明孩子的家长的愉悦互相平衡。当不开心的家长做出种种努力使孩子与标准一致时，产生的危害将无法弥补。

阿诺德·杰塞(Arnold Gesell)博士和他的方法并没有反映所有内容。错误出在从研究者经过耸人听闻或消息不灵通的作者向读者传递的筛选过程中，而读者又没有发现这个过程中所遗漏的数据。如果能给“准则”或者平均数加上反映全体范围的指标。那么，许多误会将消除。当发现自己的孩子属于正常范畴时，父母将排除由微小和无意义的差异引起的担心。

无论如何几乎没有人是正常的,就如同抛 100 次硬币，几乎很难出现 50 个正面和 50 个反面的情况。

将“正常的”与“期望的”混为一谈导致事情变得更糟。杰塞博士仅仅描述了一些通过观察得到的事实，这使得那些阅读书籍和文章的父母错下结论：晚一天或晚一个月学会走路的孩子是低能儿。

对艾尔弗雷德·金西博士的著名报告(几乎没有人认真读过)所做的许多愚蠢批评正是将正常误解为好的、对的、应该如此的等价物。金西博士被指责有教唆年轻人的嫌疑，因为他向他们灌输某些观念，特别是将许多普遍存

在却未经认可的性行为称为正常。实际上，金西博士只是指出他发现这种行为很普遍，而这便是正常所指，他并没有为这些行为贴上许可的标记，这些行为是否道德也并不属于他的研究范围。只是他恰巧涉及了一个使许多人头疼的话题——涉及这样一个高度敏感的话题却不迅速表明你支持或反对的态度看来是十分危险的。

这些没有透露的数据其欺骗性在于人们经常忽略了它们的不存在，这当然也是使用这些数据的人获取成功的奥秘。现在老练的报界批评家正哀叹缺乏严谨的老式工作做派，并严厉地批评“华盛顿的椅子记者”——那些仅靠政府新闻混日子的人。从新闻杂志《两星期》(Fortnight)的“新的工业进步”栏目中挑选一条新闻；“来自华盛顿的消息：一种能提高钢材硬度两倍的冷回火新方法产生了。”从中可一窥报界的无所作为。

在你认真研究这到底意味着什么之前，该新闻听上去似乎让人感到钢铁的质量真是有了长足的进步，但实际上它却像水银球一样令人难以捉摸。是否这种新的方法让所有种类的钢材硬度达到未处理前的 3 倍？抑或它产生了硬度是以前任何一种钢材 3 倍的新钢材？新方法采取了哪些措施？看上去记者传递了一些他自己也不了解的信息，而你正期待不批判地去读它们，因为仿佛有种学到某些东西的快乐幻觉。这容易使人回忆起关于老师上课的古老定义：这是一个将教师书中的内容在没有经过双方大脑思考的情况下转化成学生笔记的过程。

几分钟之前，当我在《时代》杂志查阅有关金西博士的内容时，我看到了另一个只要看第二眼就能够否定的声

明。这是 1948 年一些电力公司联合推出的广告：“今天，超过四分之三的美国农场接上了电……”听上去真不错，这些电力公司真是尽职尽责。当然，如果你是爱挑剔的人，那么，你还可以这样解释：“将近四分之一的美国农场还没接上电。”但真正的把戏却并不在此，而在于“接上”这个词。用了这个词，电力公司可以把事情描述成他们所希望的任何效果。很明显，“接上”并不意味着所有这些农场已接通了电，否则，电力公司一定会用“使用上”一词。据我所知，他们的“接上”只能说明电线从那些农场经过或铺设在农场几十或者上百英里范围之内。



销售地区距铁路、公路、海港或者可通航的水道25英里内
(狗拉雪橇的道路未包括在内)

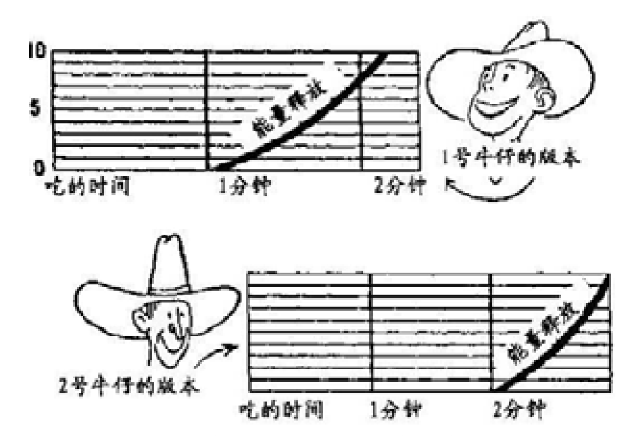
让我引用刊登在 1952 年《矿工》(Collier’s)上的一篇文章的题目——《现在就来预测孩子未来长多高》

(You Can Tell Now How Tall Your Child Will Grow)。两张表很显眼地附在正文中，一张适用子男孩，另一张适用于女孩，这两张表给出了每个年龄阶段孩于的身高与最终身高的比例。“预测孩子长大后的身高，”标题如是写道。“只需要利用现有的身高再查表中的比例即可。”

可笑的是，只要你继续看下去，便会发现文章本身就指出了这些表格的致命缺点。不是所有孩子都具有完全相同的生长方式。有的一开始长得很慢，却突然长高；有的开始长得很高。然后速度变缓；还有的人在整个成长过程中，保持着相对稳定的速度。这两张表，正如你所疑惑的，是在进行了大量测量之后所取的平均数。对于随机抽取的 100 名年轻人，他们未来的总身高或者平均身高是可以用这两张表进行预测的，但是家长感兴趣的只是一个孩于的高度，对于个体，这两张表是无价值的。如果真想预测孩于未来的身高，那么了解父母及祖父祖母的高度或许更有用——尽管这种方法同上述表格一样是不科学、不精确的，但结果的准确性至少相当。



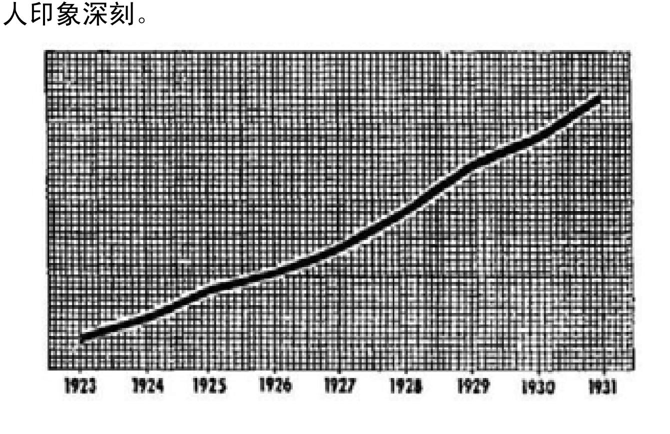
我很高兴地指出，我 14 岁进高中军训时站在最小班级的后排，利用当时记录的身高做一个预测，最终我会有大约 5 英尺 8 英寸的身高，而我现在已经是 5 英尺 11 英寸了。在人的身高中 3 英寸的差距应该能够说明该预测不太准确。

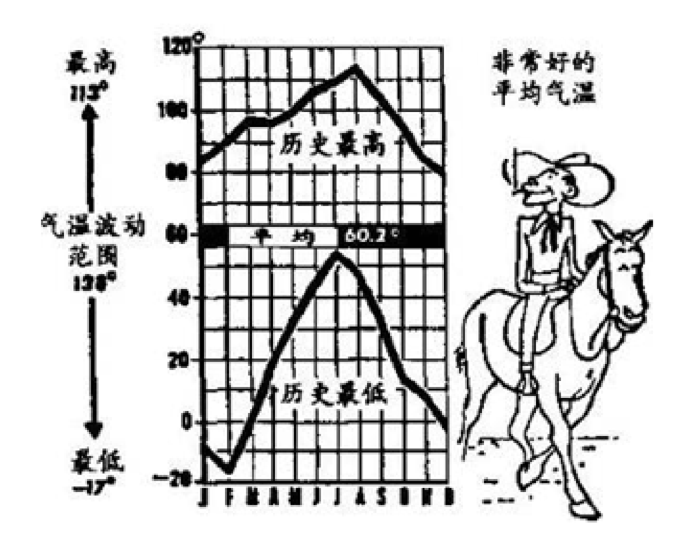


在我面前是两盒葡萄坚果饼干盒的包装纸。正如说明书上注明的，它们是不同的版本。

它们一个引用了双枪皮特(Pete)的赞语，而另一个写道：“如果你想和哈比(Hoppy)一样……你就得像哈比一样吃。”它们都提供了图表(“科学家证明是真实的！”)来说明这些薄饼“在 2 分钟之内开始提供能量！”一张图被大量文字所包围，其纵轴标有数据，另一张图省略了这些数据。既然没有关于这些数据的说明，有没有它们都一样。

两张图都有一条急剧攀升的红色曲线(代表“能量释放”)，但一条曲线开始于吃葡萄坚果饼干一分钟后，而另一条则开始于两分钟后。一条曲线的攀升速度看上去是另一条的两倍。毫无疑问，连制图者都不清楚这两张图能说明什么问题。当然，类似愚蠢的资料只会出现在青少年或者早晨疲倦不堪的父母眼前。没有人会用这种统计废话来挑衅一个著名商人的智商……难道有人会这么做？让我给你看一个广告代理机构用于宣传自己的广告(我希望它不会让人感到疑惑不解)，它刊登在《财富》(Fortune)杂志的专栏中。图中曲线意欲向人们显示这家广告公司年复一年惊人的发展趋势。但图中没有一个数字，这样一来，它既可以代表一个骇人的发展速度，每年翻番或增长几百万美金，又可以意味着在年十亿总收入的基础上，增加一美元或两美元相对稳定的蛇状爬行。但仅从图上看，其发展速度让人印象深刻。





当一个平均数、一张图表或者某种趋势遗漏了这些重要的数据，请对它们保留一些怀疑。否则，你会和一个仅根据平均气温选择野营地点的人一样盲目。也许你会认为 61 华氏度是个不错的年平均气温，而在加利福尼亚州，如果你仅根据平均气温进行选择，而忽略了气温波动范围，你可能会选择内陆沙漠或远离南海岸线的圣·尼古拉斯群岛，那么，你不是被烤焦就是被冻僵。因为，圣·尼古拉

斯群岛气温的波动范围是 47～87 华氏度，而沙漠的气温波动范围是 15～104 华氏度。

根据以往 60 年的记录，俄克拉荷马城具有十分相似的平均温度——60.2 华氏度。但是正如下图所示，这个舒适凉爽的数字遮盖了 130 华氏度的气温波动范围。

第四章、毫无意义的工作

乔赛亚·斯坦普爵士（Sir Josiah Stamp）曾经描述过这样一个场景：伦道夫爵士（Lord Randolph）正在检查一份税收报告，私人秘书站在他的身后，当看到与去年同期相比海关税收增加了 34％时，伦道夫感到很满意。

秘书马上纠正他，指出增长率只不过才.34％。

“这有什么区别？”伦道夫爵士问道。

当听到一个数值是另一个的 100 倍时，伦道夫说：“我以前经常可以看到位于数字前的小数点，但是直到现在我才真正知道它们的作用。”

在比较考试成绩时，不单单是小数点，甚至其他一些讨厌的细微差异也会突然冒出来给大家造成困扰。

如果你不介意，我们将赋予你两个孩子。如同许多受教育的孩子一样，彼德(Peter)和琳达(Linda)(我们顺便给他们起了很时髦的名字)接受了智力测试。目前任何形式的智力测试都有点原始伏都教(伏都教：又译巫毒教，源于非洲西部，是糅合祖先崇拜、万物有灵论、通灵术的原始宗教)盲目崇拜的味道，因此你费了不少口舌来打听测试的结果。这个信息是保密的，一般只有心理学家和教育者才知道，或许这样是对的。但不管怎样，你还是通过某种方式探听到：琳达的智商是 101，彼德只有 98。当然你很清楚：智商的平均数是 100，即 100 意味着“正常”。

哈，琳达是比较聪明的孩子，而且她的智商高于平均

水平，彼德则低于平均水平。对此我们先不详细评论。

任何类似的结论纯粹都是胡说。

为了澄清事实，我们首先必须注意到，无论智力测验测试什么内容，它都与我们平常意义上的智商相去甚远。它忽略了类似领导才能、创造性想像力等十分重要的素质；它没有考虑到社交判断力以及音乐、艺术或其他方面的才能；它无法测试出诸如勤劳、情感平衡等重要的人格品质。再加上，大部分学校做的智力测试都是简单低廉的类型，它们极大程度地依赖于阅读能力、测验者反映的快慢等因素，阅读速度慢的人根本没有拿高分的希望。

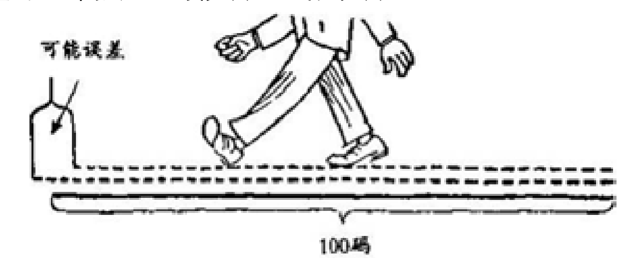
假设我们对这些都有了重新认识并一致达成共识：智力测验仅仅测验了处理那些预先准备好的抽象问题的能力，而这些能力又很难确切地进行定义。假设彼德和琳达所做的是公认最好的智力测验——修订的斯坦福—比奈测验。这个测验是独立进行并且不需要特别的阅读能力。

智力测试只是智商水平的一个抽样。与其他抽样的结果一样，代表智力水平的数据也具有统计误差，这个误差将用来衡量该数据的准确度和可信度。

智力测验类似于你估计某块地玉米质量时所做的工作。你在地里四处走动，随意地到处摘取一些玉米穗，当剥开并研究了大约 100 颗玉米穗时，你就能对整块地的质量大致有数。

当已知两块地的质量不同时，这些信息已经足够对它们的质量进行比较了。但如果接近，你还得摘取更多的玉米穗，并利用精确的质量标准来给它们划分等级。

你的样本以多大的精度代表总体是可以用数据来衡量的，那就是：可能误差和标准误差。



假设你要完成一项丈量任务，即通过步测栅栏来得到几块地的大小。你要做的第一件事自然是检查丈量体系的准确性，通过多次步量出自认为的 100 码，并发现平均而言有 3 码的误差。也就是说，一半情况下与真实 100 码的误差在 3 码之内，而另一半情况下与真实 100 码的误差在 3 码之上。

这样一来，你的可能误差就是 100 码将产生 3 码的误差，用相对数表示就是 3％。从此之后，每次被你步量测出的 100 码应该被记录成 100±3 码。

（相比较而言，大多数的统计工作者更倾向于使用标准误差。2／3 的单位将落在加、减一个标准误差的范围内。而不是 1／2 的单位。而且，标准误差的计算方法相对简单。但这里，我们还是坚持使用可能误差，并将其运用到斯坦福—比奈测试中。）假设智力测验与我们假想的步量任务有相同的 3％的可能误差。这与测验的好坏无关，而只是反映了测验与它所要测试的内容是否一致。这样，彼

德的智商应该是 98±3，琳达的智商则是 101±3。

这只不过意味着彼德的智商以相等的机会落在 95～101 中任何一点上，并且大于或者小于 98 的可能性完全一样。同样，琳达的智商落在 98～104 范围内的可能性也不过 50％。这样，你会很快发现彼德的智商有 1／4 的可能性将超过 101，而琳达则将有 1／4 的可能性低于 98。这样看来，彼德的智商并不低于而是高于琳达的智商，它们之间还有 3 分的差距。

结论是，对待智力测验以及其他抽样结果应注意它的范围。正常的智商不应该只是 100 这样一个数值，而应是一个范围，例如 90～110。将处于这个范围的孩子与低于或高于此范围的孩子进行智商比较时会得出一些有用的结论。但比较相差不大的两个数据则毫无意义。你必须记住这个加减号，即使(特别是)它没有明确指出。

忽视抽样所隐含的误差将导致一些愚蠢的举动。那些把读者调查奉若神明的杂志编辑，是因为他们不理解调查。对一篇有 40％的男性读者喜爱的文章与另一篇只有 35％的男性读者喜爱的文章，他们会刊载更多类似于前者的作品。

对于杂志而言，40％与 35％读者量的差别是很重要的，但抽样调查形成的差别并不一定是真实的。出于成本的考虑，实际的样本容量，特别是已经扣除了那些从来不读该杂志的人后，也许只有几百人。对于一本女性杂志，样本中的男性读者会很少。当这些人又根据他们的回答“全部读了”，“读了大部分”，“读了一部分”以及“没看”这篇文章而被划分成 4 组后，35％的男性读者也许只是几个人。

隐藏在这个看似令人注目的数据背后的误差也许很大，依靠它抉择的编辑并没能抓住一根救命的稻草。

人们有时会为了一个数学上可论证却小得没有意义的差别费尽力气。这种行为是对这句古训的藐视：只有当差别有意义时才能称之为差别。在下面的案例中，“老黄金”(Old Gold)香烟公司利用了一个毫无价值的结论赚了一笔。

故事起源于《文学文摘》某编辑的一个无辜想法。当他发现虽然自己抽烟却对这些香烟一无所知时，他的杂志开始行动起来，聘请一些实验室人员分析了不同品牌香烟的烟雾，并刊登了最终结果，列出每种品牌香烟的烟雾中所含尼古丁以及其他有害物质的含量。在详尽的数据支撑下，该杂志声明：所有品牌的香烟是一样的，当你吸烟时，品牌没有任何差异。

也许你会认为，对于香烟生产厂商和那些为香烟杜撰崭新广告词的家伙而言，这是一个有力的打击，它还将引爆关于某些品牌的香烟有助于平缓喉痛和对胸部有益的广告索赔案。

但有的人有了其他发现，在一长串具有相同危害物质的品牌名单上，总有一个排在最后，而那就是“老黄金”牌香烟。于是，电报漫天飞舞，大幅广告以最大的字体刊登在报纸上，其标题和副本仅仅提到，由一家国家级杂志主持的实验证明“老黄金”香烟在不良物质，以及尼古丁含量方面“排名最后”！甚至连所有可以说明差异的数据和文字都被省略了。

最后，“老黄金”公司被通知“中止并停止”使用这

个具有误导性的广告。但这并不要紧，它们早已获得了足够的好处。正像《纽约客》(New Yorker)说的，都是广告人搞的鬼。

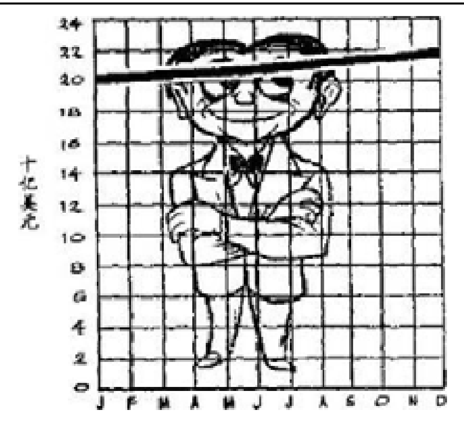
第五章、惊人的统计图形

人们对于数字似乎有一种天生的畏惧。当汉普蒂·汤普蒂(Humpty Dumpty)充满自信地告诉爱丽斯(Alice)，他能熟练地驾驭文字时，恐怕没有多少人能将同样的自信延伸到对数字的掌握上。也许中小学的算术给我们的心灵造成了创伤。

不管什么原因，当作者渴望自己的书有人读。广告商希望自己的广告能促进商品的销售。出版商想向人们炫耀自己的书或杂志是畅销书籍时，数字便产生了真正的问题。当表格中的数字是禁用的，而文字又不能达到很好的效果时，人们就自然会想到另一种解决问题的办法——画图。

最简单的统计图形是直线类图形。在显示趋势时，直线图形非常有用。而对于趋势，人们总是津津乐道于发现它、分析它，甚至预测它。下面，我们将用图形来显示国民收入怎样在一年内实现了 10％的增长。

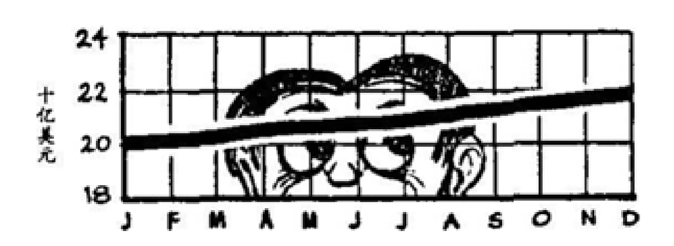
首先，在纸上用相互垂直的直线画出许多小方格。然后，在横轴的底部注明月份，在纵轴旁由下往上地标上数字“0，2，4……”并注明数字的单位“十亿美元”。在图中点出每个月的国民收入，再用直线将这些点连接起来，你将获得类似下面的图形。



这个图清楚地显示了一年来的变化，而且变化是逐月反映出来的。然而，画图者很快就会发现，由于图形的纵轴从“0”开始，并且整个图形都是按照比例绘制的，虽然直线的确上升了 10％——一个了不起的上涨趋势，但视觉效果却并不振奋人心。

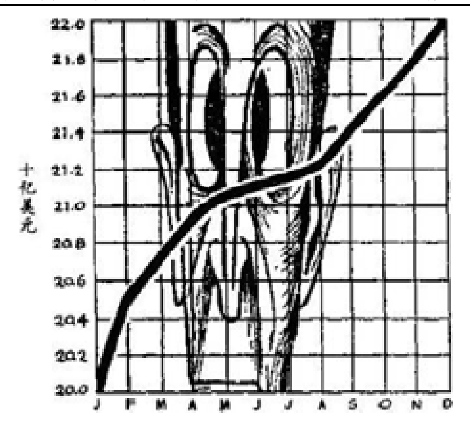
如果体所要做的仅仅是传递信息，那么目的已经达到了。但假如你希望利用图形赢得一场争论、让读者大吃一惊、促使某项行动，或者向他人推销货物等，它就缺乏渲染的效果了。试试把图形的底部抹去，不是更好吗？（如果某个爱挑剔的家伙反对这种有误导的变动，你可以理直气壮地指出，这样做节省纸张。）数据是相同的，图形也相同，除了它能给人留下印象外没有进行任何伪造。但一个粗心的读者现在所看到的，是一条在 12 个月内几乎快

升至图表顶端的直线，这仅仅是因为图表的大部分都已被抹去。就像语法课上遇到的省略句，这种图形也能被“理解”。但是，眼睛却不能“理解”被抹去的部分，这才导致微小的上升最终变成了惊人的增长。



既然已经开始行骗，那么就别急着金盆洗手。你还有比之更好的方法，它能将朴实的 10％变得比 100％的增长率更令人欢欣鼓舞。只需要稍微改变横轴与纵轴刻度的比例关系，将纵轴的每一个刻度缩减为原来的 l／10 即可。没有人规定不能这样做，而这会产生一条更完美的折线——绝对的视觉冲击。

令人震惊！不是吗？任何看到这幅图的人都会强烈地感觉到在国家的各条经济命脉上正快速地积累大量的财富。这相当于将“国民收入增长了 10 个百分点”改写成“国民收人惊人地攀升了 10 个百分点”。显然，图形比文字更有效，因为它不存在任何形容词和副词来破坏图形的客观性所形成的幻觉，而且谁也无法指责你。

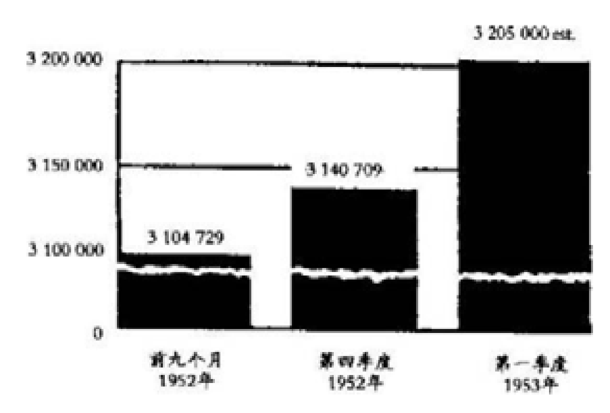


　　你还会发现一些知名的或至少是受人尊敬的同伴。《新闻周刊》(News Week)杂志在表现 1951 年“股指创了 21 年新高”时也采用了同样的做法，他们将图形底部的 80 个刻度截去。一家哥伦比亚煤气公司于 1952 年在《时代》杂志上刊登了“来自本公司最新年报”中的一张图表。如果仔细阅读图表中的那些小数字并进行分析，你会发现 10 年来生活指数上升了 60％，而汽油成本下降了 4 个百分点。这是一个中肯的描述，但对于这家公司而言显然不够理想，于是他们将图形顶部截至 90％的刻度(没有断层也没有任何文字说明进行提示)，以至于单纯通过观察得出的结论是：生活指数是原来的 3 倍，而汽油成本则下降了 1／3。

　　为了将公众反对提高工资的意见罗列出来，钢铁公司

使用了类似的具有误导性的图形。然而，它并不是什么新方法，而且它的不完善性很早就被揭露出来——并不仅仅刊登在统计专业者使用的技术刊物中。1938 年，《丹斯评论》(Dun's Review)的某个编辑曾从一则鼓吹华盛顿广告业的广告中摘录了一张统计图，图的标题是：“政府支出急剧上升!”图形中的折线与标题中的感叹号遥相呼应，尽管折线后的数据并不令人乐观。下左图那根折线从图形底部激增至顶端，将原本从 19500000 美元到 20200000 美元仅仅是 4％的增长，描绘得仿佛是 400％。这本杂志还在该图旁边给出了根据相同数据绘制的另一版本——一张客观反映 4％增长的红色折线图，标题是;“政府支出保持稳定。”

《矿工》杂志在报纸广告的柱状图中也使用了相似的技巧，但特别用文字指出：柱状图的中部被省略了。



第六章、平面图形

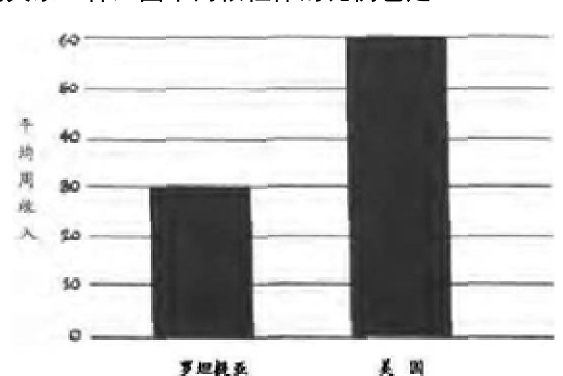
大约 10 年前，你能经常听到“小人物”这种说法，他其实指的就是我们自己。当人们觉得这个词过于卑微后，就改用“普通人”，但很快这个提法也被遗忘了。但是“小人物”仍然跟随着我们，他扮演着统计图中的角色。

用一个小人来表示成千上万的人，一个钱袋或一堆硬币表示成千上万的钱，一片牛肉来表示明年牛肉的供应量，这些都是形象化的图形。由于十分吸引人，它可以作为一种有用的工具，但同时它也能摇身一变，成为一个老练、狡猾而且成功的骗子。

形象化图形的前身是普通的柱状图，在比较两种成两种以上事物时，柱状图是一种描绘数量的便捷常用的方法。但柱状图也具有欺骗性。在代表一种事物时，柱体的长度和宽度被随意地改动；在代表三维物体时，物体的体积又不容易进行比较，因此我们应该对这种图形保留一些怀疑。一个被截短的柱状图与我们曾讨论过的被截短的折线图实乃一丘之貉。柱状图通常出现在地理书籍、公司声明以及新闻杂志中，也是因为它的“迷人”特性。

如果我想对两个数据——例如美国与罗坦提亚(Rotunda)木匠的平均周工资——进行比较，假设它们分别为 60 美元和 30 美元。为了吸引你的注意，我不会满足于仅仅将数字打印出来，而是画了柱状图。(顺便提一句，如果 60 美元与去年你为了给游廊做一个新扶手而支付给木匠的大笔工钱不一致的话，请记住木匠并不是每周都有

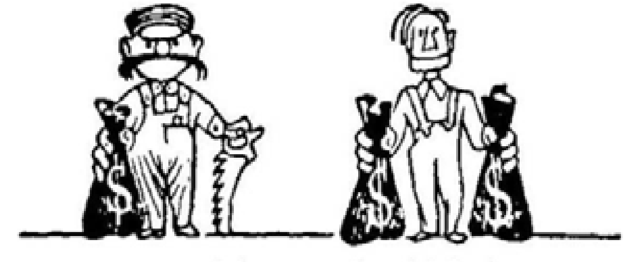
这么好的收入。而且，毕竟这里我也没有指明使用了哪种平均数，因此它不值得你去争论。瞧瞧，只要你省略了其他一些信息，那么伪装哪怕是最声名狼藉的统计资料又是多么容易呀！也许，你会怀疑我为了举例而编造了这个数据，但我敢肯定，如果这里用的是 59.83 美元，你就根本不会有这个念头。)这就是柱状图，其纵轴表示木匠的平均周收入。这是张清楚且忠于事实的图。同收入 1：2 的比例关系一样，图中两根柱体的比例也是 l：2。



但这张图并不形象，不是吗？我可以用比柱体看上去更像钱的东西——钱袋来进行改善。1 个钱袋表示罗坦提亚木匠可怜的收入，两个钱袋表示美国木匠的收入。

或者 3 个表示罗坦提业水匠的每周收入。而用 6 个表示美国木匠的每周收入。不管怎样，这样的图仍沿袭了真实可信的作风。它不会利用你匆忙的一瞥而欺骗你。这是

一个老实人的作品。



罗坦提亚木匠每周的收入　　美国木匠每周的收入

如果仅仅为了交流，这么做已经足够了。但我还有另外的目的，我希望说明美国木匠比罗坦提亚木匠的境况好得多，30 与 60 之间的差距越被夸大就对我越有利。换句话说(当然这并不是我真正的意图)，我希望你能从中推断出什么，能留下一个夸张的印象，而我又不会因此惹上麻烦。下而介绍一种方法……这种方法几乎每天都被用来愚弄人。我随手画一个钱袋表示罗坦提亚木匠的 30 美元，然后再画一个高两倍的钱袋来代表美国的60美元。还是1：2 的比例，对吗？但现在却达到了我所追求的直观感受——美国木匠的工资使罗坦提亚木匠相形见绌。



罗姆提亚木匠每周的收入



美国木匠每周的收入

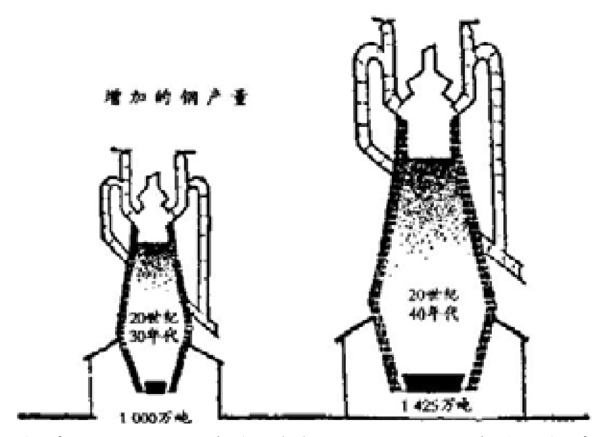
奥妙在于，既然第二个袋子比第一个高一倍，那么，应该同样宽一倍，占用纸张的面积就不是 2 倍而是 4 倍。数字全是 2：1，但视觉效果——大多数情况下它起着决定性的作用，却是 4：1。更糟的是，既然实际事物往往是三维的，那么，第二个袋子还应该比第一个袋子厚一倍，因为几何知识告诉我们，相似物体的长、宽、高保持相同的变化。于是，2 乘 2 乘 2 等于 8，如果一个钱袋里有 30 美元，另一个应该是它的 8 倍，即 240 美元。

这就是我的富有创造性图形的威力！在说“2 倍”的同时，最终给人留下的印象却是令人震惊的 8：1。

你却很难指责我有犯罪的意图，因为我不过在跟风随大流。《新闻周刊》杂志也使用过类似的钱袋。

美国钢铁协会(The American Iron and Steel Insti-

tute)也这么做了，不过用的是一对爆破熔炉。他们希望通过从 30 年代到 40 年代大幅度提升的钢铁产量，来证明在没有政府干预的情况下，该行业也能很好地自我运作。原理本身比表现原理的手段更值得推崇。



代表 20 世纪 30 年代增产 1000 万吨的熔炉是代表 40 年代增产 1425 万吨熔炉高度的 2／3。而眼睛看到的两个熔炉，后者却是前者的 3 倍。嘴上说的是 1.5 倍，眼里看到的却是 3 倍……这是一维图形的佳绩。

美国钢铁协会的这个艺术品还有其他有趣之处。从水平上看，第二个熔炉似乎宽了些，其宽度与邻居的比例失调，同时，熔炉内的黑色条块，代表着熔化的铁，其长度看上去是 10 年前的 2.5 倍。于是，50％的增长被画成了 150％的增长，视觉效果又使其变成了 1500％的增长——

算术简直变成了魔术。

（将这种彩色插图看成是被截短的折线图的优良标本，似乎过于苛刻。一个通过将图形下部分截去而夸大了人均钢产量的曲线，不仅节省了纸张，同时使产出增长率攀升了两倍。）

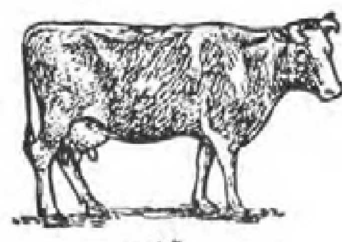
有些例子可以借口制图者水平实在太差，但如果发现出纳老是少找钱，而所有的差错都对出纳有利时，那你就禁不住要怀疑了。

《新闻周刊》曾经利用图形来证明“美国人长寿了”的说法，图形中的两个男人图案，一个代表 1879～1889 年间人们的平均预期寿命 34 岁，另一个代表现代人的平均预期寿命 68.2 岁。又是这个鬼把戏：一个人是另一个的两倍高，按此推理，体积或重量的比例应该是 8：1。通过夸大事实来达到更好的效果，我把它称为“新闻炒作”的一种形式，同期杂志上还有一张被截短的折线图，十分惊人。

在图中任意改变事物的大小还有一个弊端，1860 年美国只有 800 万头奶牛，而到了 1936 年数量增加到了 2500 万头。为了显示产量的增长，画两头奶牛，一头是另一头的 3 倍高。当然，这幅图也会以我们探讨过的方式夸大人们的印象。但更危险的是，通过对内容的快速浏览，粗心读者将极容易产生错误的结论：现在的牛比以前的牛要大得多。

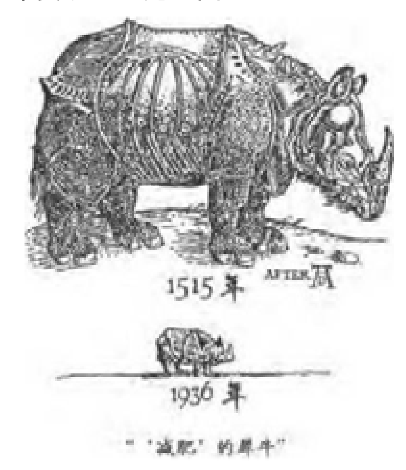
“变大的奶牛”





1860年　　1936年

同样的技巧也可以运用在犀牛的个数上。当然，这并不意味着犀牛变小了，见下图。



“‘减肥’的犀牛”

第七章、相匹配的资料

如果你想证明某事，却发现没有能力办到，那么试着解释其他事情并假装它们是一回事。在统计资料与人类思维冲撞所引起的耀眼光芒中，几乎没有人会发现它们的区别。不相匹配的数据是保证你处在有利位置上的武器，而且通常如此。

你无法证明你的秘方能够治疗感冒，但你却可以在报纸上(用大字标题)刊登一篇赌咒式的实验室报告，在 11 秒内仅仅半盎司的该药剂量就杀死了试管中 31108 个细菌。你必须确保这家实验室是有名的，或者有一个让人印象深刻的名字，然后将这个报告全文复印下来，并在旁边放上一个白大褂医生的照片。

但记住千万别提这里面的小把戏，而且你也没有责任指出，在试管中很有效的抗菌剂到了人的咽喉中就不能发挥作用，特别是为了防止该药灼烧喉咙而按照说明书进行稀释后。也不要为报道杀死了哪些细菌而迷惑。谁会知道哪种细菌引起了感冒？说不定感冒根本与该细菌无关。

实际上，试管内到底有哪些细菌与感冒的制造者之间有什么联系仍是未知数。但大家并不会提出质疑，特别正流着鼻涕时。

也许这个例子太明显了，大家很容易识破它。然而，不匹配的资料通常不会以这种面目出现在广告中。下面，让我们欣赏一个更高明的手法。

假设在美国种族歧视情绪上升的时期内，你受雇去

“证明”相反的结论。这并不是件难事。进行一项民意测验，或者更好的做法是聘请一家声名显赫的调查机构来替自己进行调查。对总体中通常的横截面样本进行提问，对“是否认为黑人与白人有相同的就业机会”这一问题间断地进行调查，以确保你能得到一个很像样的报告。

民意研究学会普林斯顿办事处(Princeton’s Office of Public Opinion Research)曾经测试过同样的问题。得到的结论是极其有趣的：一些结论，特别是民意测验的结论，通常与实际情况并不一致。被调查者在被询问关于就业机会问题的同时，还回答了其他的问题。

这些问题旨在发现被调查者真实的种族歧视态度。结果证明，大部分特别歧视黑人的人对就业机会问题的回答是“是”。(实际上那些同情黑人的人中三分之二的人并不认为黑人与白人有相同的机会，面具有种族歧视倾向的人中三分之二的人认为就业机会相等。)很明显，这个调查并不能使你对黑人的就业情况有很好的了解，而只能使你从中得到有关人们种族态度的一些有趣结论。

于是，在调查期间，如果种族歧视情绪抬头，就将有更多的人认为黑人与白人就业机会相等。于是就可宣布调查结果：黑人一直受到很公平的对待。

在巧妙运用不相匹配的数据后，你得到了自己想要的结论。如同上例，实际情况越糟，你的结论却越让人欣慰。

再看看这个例子：“对著名内科医生香烟品牌调查的结果：27％的被调查者选择了喉宝(Throaties)——该比例高于其他品牌。”这个数据本身也许就是虚构的，但不用管它。对这种不相关的数据，惟一的回答是：“那又如

何？”是否拥有了那些对医务职业的尊敬，医生就能比其他人掌握更多关于香烟品牌的资料？是否自身的知识能够促使他们选择危害性最小的品牌？当然不是，也许你的医生便会第一个反驳这种观点。然而，“27％”的提法又仿佛暗示着它能解释一些问题。

把上述比例缩小 1％，并运用在榨汁机上。作为一种新发明，它的广告随处可见，“经过实验室的证明”，这种榨汁机的“榨汁功能增强了 26％”，并且得到了“某家著名家政研究机构的推荐”。

听起来的确不错。如果有这样一台功能增强 26％的榨汁机，为什么还买别的产品呢？现在，我们暂且不去深究“实验室的实验”（特别是“实验室的独立试验”）到底证明了什么。现在让我们看看根据这个数据能得到怎样的结论？功能增强了 26％，比较的对象又是什么？如果发现不过是一台老式的手摇榨汁机，它恐怕就不是大家值得掏腰包的了，也许它是市场上最差的一种榨汁机。除了令人可疑的精确外，26％是毫无意义的。

广告商并不是惟一会利用数字糊弄你的人。一篇刊登在《本周》（This Week）杂志上探讨驾驶安全的文章毫无疑问将引起你的兴趣。文章告诉你，当你“以每小时 70 英里的速度疾驶在高速公路上”，如果时间是早上 7 点，那么你生还的机会将是晚上 7 点的 4 倍。证据是，“晚上 7 点发生的灾难是早上 7 点的 4 倍”。证据基本正确，但根据这样的证据是否能得到文中的结论？晚上的车祸比早上多，仅仅因为晚上有更多的车和人在高速公路上。单独一个驾驶员在晚上也许会比较危险，但上述的数据却不足

以证明这一点。

以同样荒谬的逻辑继续推理下去的话，你还可以证明天气晴朗时驾车比有雾时更危险。因为晴天比雾天多，所以天气晴朗时会有更多的交通意外。但只要运用常识，你我都能知道雾会使驾车变得危险。

在运用交通意外事故的数据时，如果不牢记它们是极其不匹配的数据，那么，无论哪种交通手段的事故记录都会将你吓个半死。

“去年飞机失事造成的人员死亡比 1910 年多”，这是否意味着现在乘飞机要比过去危险？认为更危险的说法是不合理的，因为选择飞机作为交通工具的人已经是以前的几百倍了。据报道，在最近的某一年中，有 4712 人死于蒸汽火车交通事故。这听起来似乎坚定了人们远离火车，或者应该自己开汽车的信念。但是，通过进一步深入调查，你将得出完全不同的结论。几乎一半的死亡者是那些驾驶汽车在十字路口与火车相撞的人，而剩下的人中大部分是无票偷乘火车的流浪汉。4712 人中仅有 132 人是火车上的乘客。而且，除非将这个数据与总旅客路程数相结合，否则 132 人也不能直接用于比较。

如果你即将出门远行，这是一趟穿越美国的长途跋涉，而你又十分关心旅途的安全，那么询问去年火车、飞机和汽车谁的意外事故数多，直接比较这些数据从而判断危险性是不正确的。询问每 100 万个乘客中的遇难人数，然后再计算比例，比较这些比例，才能知道最大的危险所在。

搜集这样的资料，却把它说成是另一回事，这种挂羊头卖狗肉的行为在生活中还有许多其他的形式。最普遍的

做法是将看上去极像，而完全不同的两件事混淆在一起。假设你是某公司的人事部经理，而公司与工会发生了摩擦，于是你着手进行一项“调查”，以统计多少职员对工会不满。除非工会是由人人心中向善的天使组成的团体，否则，在你绝对真实的询问和记录下，结果往往是大部分职员对工会有怨言。于是，你可以这样公布结果；“大多数——约 78％的职员反对工会。”你所做的只是将对工会有无关痛痒的牢骚，无实质内容的诉苦的职工统统加进来。你的结论根本不能说明什么问题，但听上去，它却支持了职工反对工会的观点。

这的确能起到一些作用。工会也可以采用相同的技巧进行回击。他们会很乐意“证明”，实际上所有的工人都反对公司目前的经营方式。

如果你想继续搜寻这种不相匹配的数据，那么在公司的财务公报中，你将大有收获。留心那些看上去大得惊人的利润或隐藏在其他名目下的利润。美国汽车工人联合会(The United Automobite Worker)的杂志《子弹》(Ammunition)描述了这种策略：

公报显示，去年该公司赢利 3500 万美元，即每销售 1 美元赢利 1.5 美分，你会觉得这家公司真是不幸。假设公司厕所的一个灯泡烧坏了，换掉它得花 30 美分，换句话说，这是 20 美元销售额的利润。

但实际上，公报中的利润仅是实际利润的一半或三分之一，没有报道的利润隐藏在贬值、特殊贬值名目下以应付将来的紧急情况。

百分数也有同样的幽默感。最近 9 个月内，通用汽车

公司有一个相对稳健的税后销售利润……12.6％。但同期该公司的投资利润率竟高达 44.8％，这到底是好还是坏呢？这完全取决于你想赢得怎样的争论。

同理，《琴师》杂志的一位读者在来信选刊专栏中将 A＆P 商店的销售净利润只有 1.1％归结为该公司的自我保护。他问道：“是否每家美国私营公司都害怕被指责为奸商……因为一年中每 1000 美元的投资仅实现了比 10 美元多一点的利润？”

将该比率与大家熟悉的 FHA 抵押利率或者银行贷款利率，即 4％～6％或者更高的利率相比，1.1％听上去的确太少了。这是否意味着 A＆P 商店退出百货行业并将它的资金存入银行，靠利率过活会更好些？

诡计在于每年的投资回收率与销售总收益不是一回事。就像《琴师》杂志后来某期中所刊登的另一位读者的回答一样，“如果我每天早上以 99 美分购进一件商品，并在中午以 1 美元卖出，那我只获得 1％的收益，但是全年我却获得了投资额的 365％。”

在描述同一个数据时有不同的方法。比如说，你可以将相同的事情表述为 1％的销售利润率；15％的投资回收率；1000 万美元的利润；利润上升 40％(与 1935～1939 年的平均水平相比)；或者与去年相比下降了 60％。选择一个目前最有利于你的说法，而且读到这个数据的人中，极少有人会对它的真实性表示怀疑。

并不是所有不相匹配的数据都来自于故意的欺骗。许多统计资料，包括那些对人们十分重要的医学资料，由于与原始数据不符的报道而被扭曲。在些棘手的问题上，例

如流产、非法出生、梅毒，存在着十分惊人的矛盾数据。如果你关心近期流感和肺炎的数据，你就会得出一个奇怪的结论：这些疾病几乎都只出现在南方的 3 个州，它们占了有记录病例的 80％。而实际上，真正的原因是目前只有这 3 个州仍保留着对此类疾病的记录，其他地区已废除了这一做法。

一些有关疟疾的数据也是无意义的。在美国南部地区。1940 年以前有成千上万这样的病例，而今天只有极少例，这似乎表明对于疟疾的治疗在近几年发生了有益并且巨大的进步。但实际上，目前只有在确诊后才进行记录；而在以前，疟疾是南方许多人用以表示感冒或者着凉的一句方言。

在美国与西班牙交战期间，美国海军的死亡率是 9‰，而同时期纽约市居民的死亡率是 16‰。后来海军征兵人员就用这些数据来证明参军更安全。如果假定这些数据是正确的，那么促使这种差异产生的真正原因是什么？海军征兵人员根据两个数据的差异得出的结论是否正确？

这两组对象是不可比的。海军主要由那些体格健壮的年轻人组成，而城市居民包括婴儿、老人、病人，他们无论在哪儿都有较高的死亡率。这些数据根本不能说明符合参军标准的人在海军会比在其他地方有更高的存活机会，相反的结论也不能证明。

你也许听说过这个令人沮丧的新闻：“1952 年是美国医学史上的小儿麻痹症年”，这个说法基于该年有多于往年的病例。

但如果专家进一步斟酌这些数据，就会发现一些令人

鼓舞的事情。首先，1952 年有更多处于易感染期的孩子，就算发病率保持不变，也会有更多的患者。其次，人们对小儿麻痹症认识的加深，导致更多病人到医院进行诊断和轻微发病记录的增多。最后，当年有增加的经济刺激，即增加的小儿麻痹症保险以及从国家婴儿麻痹基金(The National Foundation for Infantile Paralysis)获得更多的帮助。所有这些都是对小儿麻痹症达到新高的质疑，而且当年的死亡人数也肯定了我们的怀疑。

这是一个有趣的事实，在考虑某种疾病的发病情况时，使用死亡率或者死亡人数比发病人数更合理——这是因为死亡报道和死亡记录的质量更高。在此例中，明显不相匹配的数据比表面上完全匹配的数据还要好。

在美国,每 4 年不相匹配的数据便会出现一个兴盛期，这并非因为这种数据存在自然波动的特性，而是因为每 4 年有一场竞选。1948 年 10 月共和党发表的竞选纲领完全建立在看似相互联系但实际上却毫无关联的数据之上：

1942 年杜威(Dewey)当选州长时，一些地区教师的最低年收入只有 900 美元；而今天，纽约州的教师享有全世界最高的收入水平。在杜威政府的建议下，在由杜威指定的委员会的表决下，立法机构于 1947 年从州财政盈余中拨出 3200 万美元直接用于提高教师收入水平，这使得纽约市教师最低收入水平提高到 2500～5323 美元之间波动。

也许，杜威先生想借此表明自己是教师的朋友，但是这些数据并不能证明这一点。这里使用了前后比较的老把戏，一些没有指明的因素加入到过程中，导致前后并不一致。以前只有 900 美元，而现在是 2500～5325 美元，的

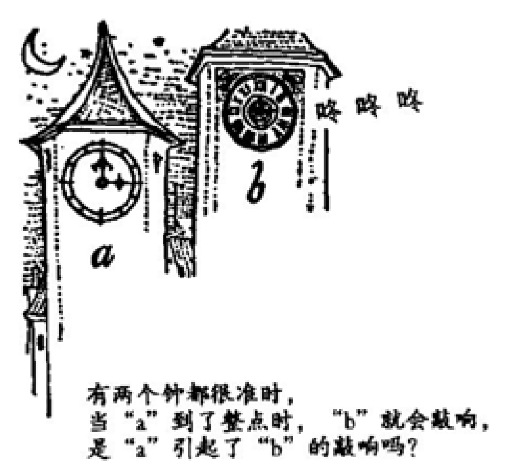
确有了长足的进步。但实际上，前者是该州乡村地区的最低收入，而后者仅仅是纽约市的最低收入水平。这些进步只能部分归功于杜威政府。

以上例子只是前后比较的统计形式，而前后图片的比较是杂志与广告中善用的绝活。同一套起居室被拍了两次，让你通过比较体会油漆的作用。但在两次拍照的间隙，增添了新家具，或者有时“以前的”照片是采光不佳的黑白照片，而“后来”的版本则是一张较大的彩照。又比如，通过照片来表现一个小女孩使用了某种品牌的洗发水后发质的改变。天哪，小姑娘的确漂亮了不少。但是，经过仔细研究，你会发现大部分的变化应归功于劝说小姑娘脸上漾起微笑的不懈努力和投射在她头发上的黑色光束。大部分赞誉应给予摄影师而不是洗发水。

第八章、相关关系与因果关系

有人曾经费尽周折以探求是否抽烟者的大学成绩比不吸烟者的差，结果的确如此。这着实让一些人高兴，从此，他们多次使用这个结论。往通往好成绩的道路上，存在着放弃抽烟的痛苦。进一步使用该结论，还有这样合理的推断：抽烟使人的头脑变笨。

我相信整个研究过程是正确进行的：样本容量足够大而且经过了认真、仔细的挑选，相关关系的确十分显著。等等。



但这是一个古老的谬误，然而它仍频繁地出现在统计

资料中，并被大量让人印象深刻的数据所伪装。这个谬误是：如果 B 紧跟着 A 出现，那么 A 一定导致 B。在上例中，当抽烟与低分同时出现时，人们得到了一个未经证实的假设，抽烟导致低分。难道就不能是相反的解释吗？也许低分促使学生不喝酒而变得爱抽烟。这种说法与前一种一样能得到证据很好的支撑。只是它不能够满足宣传人员的要求。

然而，更大的可能性是两个周素并不互为因果，而同为第三个因素的产物。是否那些不把读书当回事的爱社交的学生更爱抽烟？又或者是否可以在有人曾经建立的性格外向与成绩之间的相关关系(其相关性比成绩和智力的相关性更高)上找到线索？也许，性格外向的学生比性格内向的更爱抽烟。问题的关键在于，当有许多合理的解释时，你几乎没有权利选择对口味的解释并坚持认为它是正确的。但事实是许多人会这么做。

为了避免再次陷入到相关的谬误中，并且不再相信许多似是而非的事物，你需要对任何事物关联性的描述进行仔细的研究。所谓的“相关”，往往是通过相关系数这个精确的数据来证明两件事物具有关联关系，它可以有多种不同的类型。

一种相关是由于机缘巧合而产生的。由于机会的存在，你或许可以通过一组数据来证明一些根本不存在的结论。但换一组数据也许又无法证明。就像自称能防止蛀牙的牙膏生产厂商，你只需将对自己不利的资料扔到一边而公开你需要的结论就能达到目的。利用小样本，任意两个你能想到的事件或两组特性之间都能建立显著的相关。

联合变动的一种普通形式是存在着真实的关系，但却无法确定何为因、何为果。有时，因果可以不断地交换地位，或者可以同为因果。收入与拥有的股票之间便是这种相关。你拥有越多的钱，便能买更多的股票；同时，你手头的股票越多，又可以为你带来更多的收入。在这种情况下，简单地认为一个因素引起另一个因素是不全面的。

最具有戏剧性的相关是所有变量相互间没有任何影响，却存在着显著的相关。许多卑劣的工作都可以归为这一类。抽烟者与不好的成绩是这种例子，许多被反复引用的医学资料也是这种例子，它们只是证明了变量之间存在着相关关系，但是否存在因果关系仍旧是个未知数。下面这个真实的统计案例，也反映了这种相关。有人曾经高兴地指出：马萨诸塞州长老会的牧师收入与哈瓦那朗姆酒的价格之间存在着密切相关。

孰因孰果？换句话说，是否牧师支持了朗姆酒贸易又或者牧师从此贸易中获益？这个结论未免过于牵强，一眼便可以识破，但对于那些比这个例子更为精致的应用，我们要额外留神。在牧师和朗姆酒的例子中，人们很容易发现，如果受到第三个因素——历史性或全世界范围内物价水平上涨——的影响，那么收入和价格两个数据就都会上升。

再来看看这种说法：六月是自杀率最高的月份。那么，是否自杀导致了六月的婚礼——抑或六月的婚礼促成了被抛弃者的自杀？一个更合理的解释（同样未经证实）是：整整一个冬季，某人独自抚慰自己沮丧的心情，并满怀希望春天一切会重新开始，可是到了六月他仍没能从低落中

解脱出来，于是用自杀表示放弃。

另一个需要留意的是，超过了推断相关关系的数据范围而得出的结论。从常理来说，雨下得越多，谷物则长得越高，收成越多。雨是农民的福音。但一季暴雨则可能破坏甚至毁灭庄稼。正相关到了一定的程度后便急剧地转化为负相关。超过了一定的降雨量，雨越多，收成越少。

我们再来探讨一下教育价值量的问题。首先，假设我们已经证实了高中毕业生比中途辍学者赚的钱多，并且在大学里多学一年，其收入会有相应的提高。但我们得留意这样一个一般的结论：在学校呆的时间越长，你赚的钱就越多。注意，超过大学学历我们还没有进行证明，而原来的结论在这种情况下很可能并不适用。具有博士学位的人通常变成了大学教师，而不会成为最富有的阶层。

相关显示了一种趋势，而这种趋势通常并不是那种理想的一对一关系。平均而言，高个子男孩比矮个子男孩更重，这是个正相关。但要找到一个身高 6 英尺体重却轻于 5 英尺的人，也是一件很容易的事，因此，相关系数小于 1。负相关可以简单地描述为当一个变量增大时另一个变量有减小的趋势，物理学中，这被称为成反比关系。离一盏灯越远，你就越看不清手中的书，也就是说，当距离增大时，光线的密度将减少。这些物理学的关系一般具有确定的相关。然而，来自商业、社会学或是医学的数据却很难如此清晰。就算教育能够增加收入，但也很容易证明教育同样会导致破产。因此，请时刻记住，即使某种相关关系是真实的，并有真实的因果关系，仍旧不能凭此做出行为决策。

大量的数据被搜集起来以证明大学教育将带来的收入。许多出版的小册子利用这些数据以及多多少少建立在这些数据基础上的结论来吸引未来大学生的注意。我并不反对这种意图。相反，我十分赞成教育，特别是包含了基本统计课程的教育。这些数据充分地显示与那些未上大学的人相比，大学生将获得更多的收入，虽然有很多例外，但总体趋势是非常明显的。

数据是真实的，然而不妥的是依据这些数据和事实推断出了一个未经证实的结论。这便是相关中的谬误。结论认为，如果你(或者你的儿子、女儿)上大学，你将比用其他的方式度过这 4 年时间获得更多的收入。这个结论完全建立在同样未经证实的假设之上——既然大学生将赚更多的钱，那么，如果他们赚的钱多就是因为他们读了大学。实际上，那些能赚大钱的人就算不读大学也一样能赚大钱。有两个事实都证明了这一点。大学里大部分的学生是这样两类——聪明的和富有的。就算不读大学，聪明的孩子也有赚大钱的潜力；至于富有的……是的，钱总会生钱，不管他们是否读了大学，富人的孩于一般都不属于低收入阶层。

下面一段话摘自子一篇一问一答形式的文章。该文章刊登在《本周》杂志(《星期天》发行量极大的增刊)上。你也许和我一样觉得有趣，因为这篇文章的作者曾经写过一篇题为《流行观念：是对还是错？》的文章。

问题：上大学对你保持独身有怎样的影响？

回答：如果你是个女孩，它会提高你成为未婚女子的几率。但如果你是一个男子，结果相反——它将减少你单

身的机会。

康奈尔大学对 1500 名中年的毕业生进行了调查。他们中的男孩，93％已经成婚(同年龄段的比例为 83％)。

但是女孩中只有 65％成婚。未婚比例是同年龄段女孩的 3 倍。

当 17 岁的女孩苏西·布朗(Susie Brown)看到这段文字时，她会认为上大学将减少她未来获得新郎的概率。文章的确是这么说的，而且这些统计资料还来自于著名的高等学府。但是请注意，虽然资料来自康奈尔大学。但结论却不是康奈尔大学得出的。不过粗心的读者往往会这样认为。

同样，这里又用了真实的相关关系来支持一个未经证实的因果关系。也许，这里的因果关系正好相反，这些女孩如果不上大学也同样会保持独身，说不定数目还会更多。如果这种可能性不比作者坚持的观念更好，这只能说明：它们都只是猜想。

实际上，想保持独身的性格倾向会促使他们上大学，金西博士曾经发现性欲与教育存在着相关，这个相关受到读大学前就已定型的性格的影响。这个结论导致读大学会阻碍结婚的说法更加可疑。

于是，可以使苏西·布朗聊以自慰的是：事情并非一定如此。

一篇医学文章曾严厉警告：喝牛奶的人中癌症的发病率在上升。在新英格兰、明尼苏达州、威斯康辛州、瑞士，这些牛奶产量和消费量极大的地区，癌症有上升趋势，而

牛奶销量少的锡兰却极少发现癌症病例。更进一步的证据是，在牛奶消费量少的美国南部地区癌症病例也相对较少。文章还指出，牛奶消费量极大的英国妇女患癌症的概率是很少喝牛奶的日本妇女的 18 倍。

更深入地挖掘下去，会发现还有很多因素都可用来解释癌症发病率的提高，其中一个因素就十分具有说服力，癌症主要发生在中年或者老年人身上。而发病率高的瑞士和前面提到的那些州，其居民寿命相对较长。研究期间，英国妇女的平均寿命比日本妇女长 12 岁。

海伦·M.沃克(Helen M. Walker)教授已经通过一个有趣的例子对“一旦两个事物共同变动，他们便存在因果关系”的观点进行了批驳。在研究年龄与妇女某些生理特征的关系时，沃克测量了走路时两脚分开的角度，你将发现年纪较长的妇女两脚的角度总是比较大。你会先想到，这是否意味着脚尖朝外走路促使人变老，但这明显是无稽之谈。那么，看来是年龄的增长造成脚尖角度的增大，而且大部分妇女随着年龄增长，脚尖的角度在不断增大。

任何此类的结论都可能是错误的，而且无法得到证实。只有当对同样一些妇女或者基本上同等的群体进行一段时间的研究后才能得到合理的结论。因为这可以排除一些因素的影响，比如说，年纪大的女人在其年轻时期可能被告知应该脚尖向外走路，而年轻女人却在一个不鼓励脚尖朝外的年代学习走路。

当你发现某些人——他们往往是当事人，在胡乱使用相关性时，请注意分辨相关是否是事件变迁的产物或时代趋势的产物。在我们这个时代的任何一对数据，例如大学

生的人数、心理研究机构同房间的病人数、香烟的消费量、心脏病的发病率、X 光的使用次数、假牙的生产量、加利福尼亚学校老师的薪水、内华达州赌博的利润等，都很容易显示出正相关关系。把上面的一个事物说成是另一个事物的成果是十分可笑的。但每天都有类似的事情发生。

那些能迷惑因果关系的统计处理和具有催眠作用的数字或者小数点，比迷信好不到哪儿上，反而更具有误导性。它就像英国新赫布里底群岛土著居民的信条：身上的跳蚤会带来健康的身体。因为通过几个世纪的观察，土著居民发现健康人的身上总自一些跳蚤而身体羸弱的人通常没有跳蚤。于是他们得出结论：跳蚤使人身体健康，每个人身上都应该有跳蚤。观察是正确的，因为它经历了多年来人们随意的检验。但这并不意味着这些土著居民的结论也是正确的。

就像我们曾经提到的，比这更不充分的证据(它们经过统计的加工处理。变得凭常识都不能揭穿)制造出了大量的医学幸运儿以及刊登在医学杂志(甚至还包括专业的杂志)上的论文。更细心的观察者最终发现了新赫布里底群岛的真相：在大多数情况下，每个居民身上都有跳蚤，这是正常情况。然而，当人们发烧(说不定还是跳蚤引起的)时，随着体温上升，跳蚤不能承受高温而引起的不适，因此就会离开。这里人们完全将因果关系扭曲、颠倒，甚至混合了。

第九章、如何进行统计操纵

利用统计资料传递错误的信息而误导他人可称得上是一种操纵行为。用一个词来概括(虽然不一定很准确)，那就是统计操纵。

本书的标题与书中某些内容似乎在暗示所有这些操纵都源于故意欺骗。美国统计协会(The American statistical Association)的某分会主席为此曾经责怪过我，他说，很多情况下是出于无知而误导了他人，并不是故意欺诈。他说的也许有一些道理①，(①据说，当有大量信件需要理时，作家刘易斯·布朗费尔德(Louis BromField)有套回复那些批评信件的做法。既不做让步，又不鼓励进一步的来信。这个回答几乎能使每个人都满意。关键的句子是“你所说的也许有一些道理。”这让我想起了一位牧师，由于在集会上，他为被抱出来接受洗礼的婴儿说了祝福的话，于是获得了妈妈们的青睐。但当妈妈们试图比较他的赞美之词时，许多妈妈还能记起他说了些什么，不外乎“真不错”之类。实际上，他那恒久不变的赞美不过是：“哦!”(微笑)”真是个好宝宝，不是吗？”)但我并不能肯定后面一种说法能让统计学家觉得少受气。也许更重要的是记住这一点：许多统计资料的歪曲和被控制并不是资深统计学家所为，出自统计学家之手的完善资料也许最终会被销售人员、关系专家、记者或者广告撰稿人扭曲、夸张、简化或是刻意地进行了挑选。

在任何情况中，不管出错者是谁，谁又会相信他是因

无知而犯了错呢？报纸杂志上的虚假图形总是通过夸大事实以求造成轰动效果，而从来不会满足于轻描淡写。以我的经验，那些代表公司提出统计论据的人，几乎不会给职工或顾客提供比实际需要更多的好处。而工会，什么时候会雇用一个不称职的统计职员而使自己在劳工案件中处于劣势？

只要是单方面的错误，到底把它归咎于不称职还是偶然事故？这很难进行判断。

扭曲统计数据的最巧妙方法是利用地图。充斥在地图上的变量，往往隐藏了事实，扭曲了关系。在这方面，我认为最得意的例子是“黑色阴影”。不久前，波士顿第一国民银行(The First National Bank of Boston)印制了这张地图，随后所谓的“纳税者集团”、报纸以及《新闻周刊》杂志又将此图广为翻印。

这张地图清楚地显示，国民收入中有多大的比例被联邦政府征取并使用了。通过将密西西比州以西的其他州(除去路易斯安那州、阿肯色州和部分密苏里州)用阴影覆盖来表明联邦政府的开支与这些州的总收入相等。

它的巧妙在于所选的都是地广人稀的州。由于人口匮乏，总收入相对就少。用同样诚实(也可以说同样不诚实)的方法，这个制图者一开始也许选择的是将纽约州或新英格兰州涂黑。这是一个面积小得多的阴影，而且没有任何冲击力。数据是相同的，但两张地图给观看者留下的是完全不同的印象。没有人会选择后者。至少，我就不知道哪家实力强大的集团会对看起来比实际支出少的公共支出感兴趣。

涂黑的阴影

(西部模式)



(东部模式)



为了证明我们没有说谎，我们在上图中多涂黑了几个州

如果制图者的只的仅仅是传递信息，这很容易做到，他可以选择一些位于中部的州，这些州的面积与国土总面积的比例等于它们的总收入与国民总收入的比例。

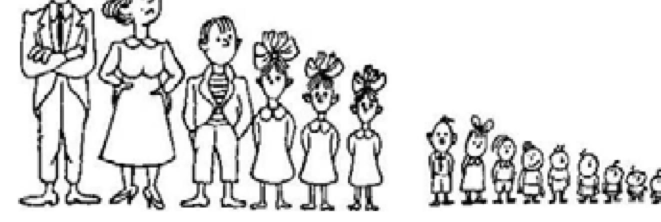
这张地图之所以臭名远扬，是因为它早已不是什么新把戏了，甚至已经成为笑柄。这家银行很早以前就曾印制了相似的地图，它是这里所看到的地图的早期版本，它用来显示 1929 年和 1937 年的联邦支出。出版不久，它们就被威廉·柯布·布林顿(Willard Cope Brinton)作为糟糕的图形范例收集在标准读物《图示》(Graphic Presentation)中。布林顿认为这种方法“扭曲了事实”。但是，这家银行仍继续画着它的地图，而《新闻周刊》杂志和其他人——他们对此图的问题应该更清楚，仍继续翻印这种地图，而且没有任何提醒和道歉。

美国家庭的平均收入有多少？在前面我们曾提到，1949 年普查局公布的数据是“普通家庭平均收入是 3100 美元”。但如果你读了由拉塞尔·塞齐基金会(The Russell Sage Foundation)发布的那篇关于“博爱的礼物”的新闻故事，你将发现，同年这个数据是惊人的 5004 美元。也许你会为大家的好收成而高兴，但同时还存有一丝疑惑——为什么这个数字与平常的观察不相符合？也许是因为观察的对象不同？

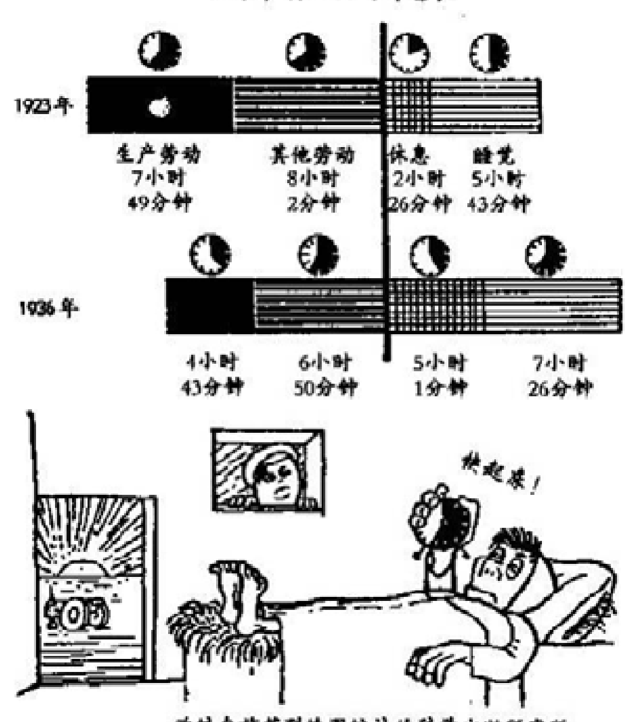
拉塞尔·塞齐基金会与普查局的数字为什么相差这么远呢？普查局用的是中位数，一种合理的计算方法，但就算拉塞尔·塞齐基金会采用的是均值，也不至于有这么大的差别。实际上，拉塞尔·塞齐基金创造了一个假想的家庭，利用它才获得了这么好的收入。具体的步骤是，他们

解释道(当被要求进行解释时)，将美国居民的总收入除以 1.49 亿人(总人数)得到人均 1251 美元，“于是”，他们补充道，“一个四口之家的平均总收人为 5004 美元。”

　　这种奇怪的算法在两方面进行了夸张，其一，它使用了均值这种平均数，面不是更具代表性的数值相对较小的中位数……这个我们在前面章节中已做了说明。其二，它假设家庭的收入和人口数成正比。现在我有 4 个孩子，真希望事实和他们描述的一样，可惜并非如此。四口之家的财富决不会是两口之家的两倍。



　　公平而言，拉塞尔·塞齐基金会并不是故意欺骗。应该说，他们主要感兴趣的是描绘一幅关于给予而不是所得的画面。这个可笑的数据仅仅是个副产品。但这并不会因此而削减它的杀伤力。这个事例也进一步说明了对那些未加解释的平均数我们根本不用太在意。



改编自前苏联绘图统计的科学出版研究所

为了使最声名狼藉的统计资料看上去更有分量、更精确，你应该考虑使用小数。询问 100 个人昨晚的睡眠时间，

进行加总，得到一个数值，比方说 783.1。任何类似的数据从一开始就不可能很精确，大部分人的回答与实际有出入，相差 15 分钟甚至更长时间都有可能，而且我们也不能保证所有的出人正负全部抵消。每个人都会有这样的经历，把 5 分钟的失眠说成是半宿的辗转反侧。不管怎样，还是先让我们继续下去。将 783.1 进行平均后宣布结论：人们平均每晚睡眠 7.831 小时。听上去，你似乎对正在讨论的话题胸有成竹。但如果你竟傻乎乎地宣布，人们平均每晚休息 7.8 小时(或者“差不多 8 小时”)，数据就失去了动人的精确性。就像这个事情本身一样，它只不过是一个可怜的近似，而且和任何人的猜想一样毫无意义。

百分数也给误解提供了肥沃的土壤。和小数一样，它也能为不确切的事物蒙上精确的面纱。美国劳工部(The United States Department of Labor)的《每月劳动回顾》(Monthly Labar Review)曾经指出，在华盛顿，某个特定月份中支付交通费用的兼职家庭帮工中，4.9％的人周收入是 18 美元。实际上，4.9％的兼职家庭帮工仅指 2 个帮工。也就是说，以上结果仅含有 41 个被调查对象。任何建立在小样本容量上的百分数都可能产生误导，直接给出具体数值的大小将更有价值。如果再将百分数表示成小数点后几位的小数形式，你可能正将天平从愚蠢的一端移向欺诈。

“现在就购买你的圣诞礼物，你将节省 100％。”这是一则广告宣传。听上去就像圣诞老人免费的馈赠，但实际上它混淆了比较的基础，相对于原来的价格，价格只缩减了 50％。虽然对于打折后的价格或新价格而言，没错，减

少量的确是 100％，但这与广告说的内容可不一样。

同样，当某鲜花联合会负责人在接受报纸采访时说，“鲜花价格比 4 个月前便宜了 100％。”他并不是想说花农已经将花免费送人了，但他的话的确是这个意思。

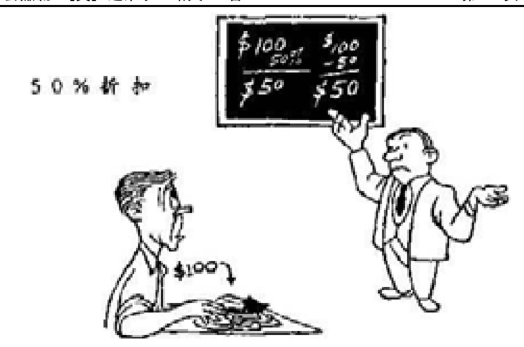
在《标准石油公司的历史》(History of the Standard Oil Company)一书中，艾达·M.塔贝尔(Ida M. Tarbell)甚至走得更远。她说，“西南部的价格下调……从 14％到 220％。”看来，为了把这些原油拖走，卖方反而得向买方支付一笔相当可观的酬金。

哥伦比亚的《快报》(Dispatch)曾宣称一家制造厂商出售某种产品能赢得 3800％的利润率，其成本是 1.75 美元而销售价格是 40 美元。在计算利润率时，你可以选择不同的计算方式(假设你有义务说出计算方法)。如果以成本为基数，结果是 2185％；以销售价格为基础，利润率仅为 95.6％。显然，哥伦比亚《快报》有一套自己的方法，并得到了十分夸张的报告数据。

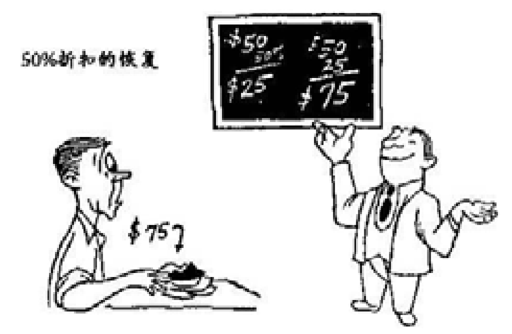
即使《纽约时报》在报道一条来自印第安纳波利斯城的美联社新闻时，也在偷换基数问题上犯了错误。

今天，经济危机已大为缓解。属于印第安纳波利斯城建筑贸易工会的管子工、泥水匠、木工、油漆工和其他工种的工人享受了 5％的工资提升，这是去年冬天工资下降 20％的四分之一补偿。

表面上很合理——但是请注意，减少是以原有工资为基数计算的，而增加却使用了较小的基数，即削减后的收入。



我们可以通过一些假设数据来进行检验。为简单起见，假设原有收入为 1 美元，减少 20％，即为 80 美分，而在 80 美分的基础上增加 5％仅提高了 4 美分，它是原来减少量 20 美分的五分之一，而非四分之一。和许多看起来似乎正确的错误一样，这个错误通过巧妙的夸张，使事情看上去更美妙了。



以上这些说明：50％的削减量需要通过提高 100％才能加以补偿。

《纽约时报》还曾经有过这样的报道，在一个财政年度内，“由于火灾，航空信笺丢失了约 4863 磅，但只占总信笺的 0.00063％。”通过该报道我们还知道，当年航空信笺约有 7715741 磅。要是保险公司按 0.00063％比例进行理赔，那么将遇到一大堆麻烦，因为根据原始数据计算的实际损失比例应该是 0.063％，为报纸所说的 100 倍。

变换基数还能产生增加折扣的幻觉。当软件批发商向你提供一份“50％折扣再打 20％折扣”的报价单时，那并不意味着 70％的折扣，实际只有 60％，因为，后面 20％的折扣是用五折后的价格计算的。

将一些看似能直接相加却不能这样操作的事情加在一起会产生大量的欺骗和隐瞒。许多孩子都曾利用这个把戏，来证明自己不用上学。

你也许还能回忆起这个小把戏。一年 365 天，减去三分之一即 122 天作为休息时间，再减去约 45 天作为一日三个小时的进餐时间，余下的 198 天中再扣除 90 天度暑假，21 天过圣诞节和万圣节。这时余下的时间连过星期六和星期天都不够。

如果个把戏太陈旧、太露骨，那么在严肃的商业活动中就不会使用它，你也许会这么想。然而，联合汽车公司的工会组织在他们的月刊杂志《子弹》中却坚持认为这种把戏至今仍在使用。

每当发生罢工时，就会有不着边际的谎言。一旦罢工

开始。商业委员便刊登广告，宣称罢工一天造成的损失是几百万美元。

这个数据是怎样得到的呢？他们将罢工工人正常生产时制造的汽车价值加起来，同时加上供应商的损失以及一切可以加上的费用，包括零售商的销售损失以及街头停车费的损失。

无独有偶，这种认为百分数能像苹果一样随意相加的观点被用来反对提高作家的收入，摘自《纽约时报》书评的下面这段文字是多么令人信服呀！

上涨的书价与作家收入之间的代沟是由生产与原材料成本的上升造成的。逐条列出如下：种植和加工成本在最近 10 年上升了 10％～12％；材料成本攀升了 6％～9％；销售及广告成本爬升了 10％以上。将所有这些加起来的话，总成本至少上升了 33％（这仅对一家大公司而言），如果只是个小出版社的话，至少上升了 40％。

实际上。如果出版一本书的每项成本开支都上升 10％左右，总成本也只爬升了相同的比例，即 10％。这种允许所有百分数直接相加的逻辑将得到各种奇谈怪论。购买 20 件物品并且发现其中每一件物品的价格相对去年上升了 5％。这是否意味着生活成本是去年的一倍（20 个 5％直接相加得到 100％）？这无疑是一派胡言。

这有些类似路边小贩的故事。当人们询问他的兔肉三明治为什么能卖到如此便宜的价格时，“哦，”他说，“我当然得掺一些马肉，但我的比例是一比一 —— 一匹马，一只兔子。”

一个工会刊物曾用卡通漫画来反对另一种没理由的直接相加。他们发现，老板将正常工作一小时支付的 1.5 美元与加班一小时支付的 2.25 美元以及加班第二个小时支付的 3 美元相加并进行平均，得到平均工资为每小时 2.25 美元。要找到一个比它更无聊的平均数还真难。

另一个困惑出现在百分比与百分点之间。如果某年的投资回报率为 3％，而第二年为 6％，你可以很谦虚地描述：增了 3 个百分点。同样，这样的描述也是允许的：增长高达 100％。要了解人们怎样混淆这两种说法，请观察民意选举的投票者。

百分位数也同样具有欺骗性。如果你被告知，在代数或其他学科方面，乔尼(Johnny)在班级的位置时，你通常会得到一个百分位数。它意味着在 100 个学生中乔尼的排名。比如说，一个有 300 名学生的班级里，前 3 名同学是第 99 个百分位数，接下来的 3 名是第 98 个，以此类推。百分位数的一个奇怪特点是第 99 个百分位数上的学生和第 90 个百分位数上学生的差异，也许与第 60 与第 40 个的差异相等。这是因为许多指标都有聚集在平均数周围的特点。与我们已经讨论过的“正态分布”的性质相同。

专业统计者的斗争不断升级，甚至最外行的人都会感到其中有诈。当善于操纵统计的人闹翻时，老实人可以稍作喘息。钢铁行业委员会(The Steel Industry Board)曾经指出了钢铁公司和工会都热衷的一些鬼把戏。为了表明 1948 年钢铁公司的经营业绩不错(以证明钢铁公司有能力支付更高的工资)，工会将该年钢铁产量与 1939 年——当年具有特别低的产量，进行对比。而钢铁公司，在这场欺

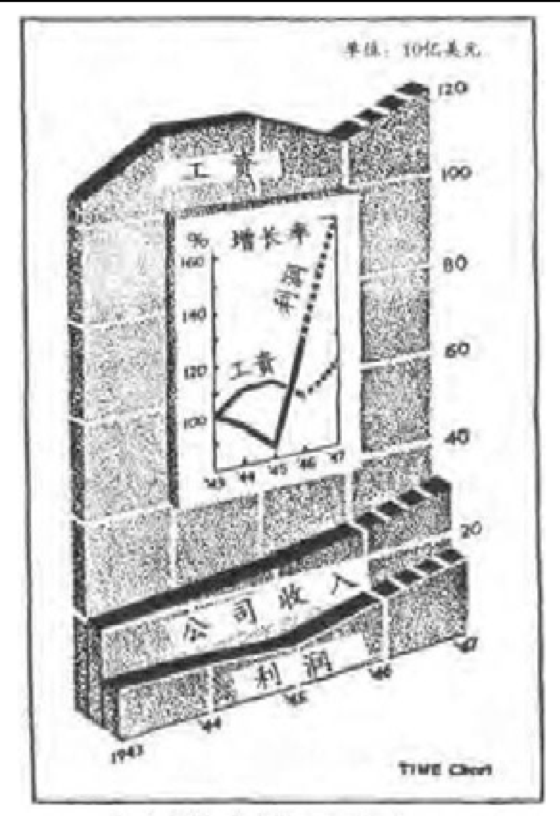
骗的竞赛中并没有被打败，他们坚持以职工获得的收入而不是以平均每小时收入进行比较。这是因为以前有许多兼职工人，即使工资率没有变化，工人的工资也一定会增加。

以擅长图表而闻名的《时代》杂志，曾经刊出了一张图，它是一个十分有趣的例子，说明统计学家能够像魔术师那样从同一个口袋中变出他所需要的任何东西。当面临着从两种同样有效的方法中进行选择，一种有利于管理层，另一种有利于劳动者，《时代》则同时用上了两种方法。这是一张由两幅图叠加而成的图形，它们都采用了相同的数据。

一张图以 10 亿美元为单位画出工资和利润，两者都呈上升趋势，而且升幅大致相同，由于工资大约是利润的 6 倍，所以看起来，巨大的通胀压力主要来自工资。

叠加图的另一部分用增长率来描述变化。工资增长率的折线相对平坦。而利润增长率的折线向上急剧攀升。从图中可以推断出，产生通胀压力的罪魁祸首主要是利润。

你可以自己选择其中一个结论。但是，如果你能看出无论哪张图都是客观真实，都无法成为指控的对象，那就更好了。有时指出一个处在争论之中的命题是否正解并不像看上去的只需一开一关那样简单，还得费些功夫。

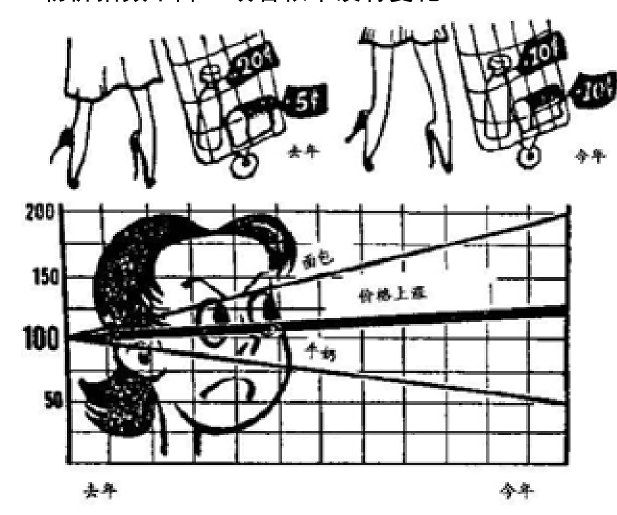


在《时代》杂志的同意下经过
重新绘制并作为一个说实话的例子。

既然工资率会受到指数的影响，那么，对数以百万计的人而言，它就是十分重要的数据。然而，值得注意的是，怎样做才能让它和着某些人的拍子翩翩起舞呢？

举一个最简单的例子，让我们假设去年一夸脱牛奶值 20 美分，一条面包 5 美分。今年牛奶的价格降至 10 美分，而面包的价格升至 10 美分。现在你想证明什么呢？物价指数上升？

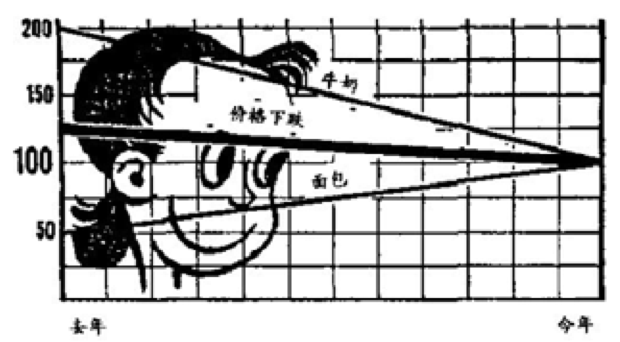
物价指数下降？或者根本没有变化？



首先选择去年作为基期，也就是说，以去年的价格为 100％。既然牛奶的价格降了一半(即 50％)，并且面包的

价格是去年的 2 倍(即 200％)，将 50％与 200％进行平均得 125％，与去年相比，今年的价格上涨了 25％。

用另一种方法试试，以今年的价格为基期。去年牛奶的价格是今年的 200％，而面包的价格是今年的 50％，平均数又是 125％,也就是说,去年的价格比今年的高 25％,今年的价格下降了。



如果你想证明价格没有发生变化，试试使用几何平均数，这时你可以随意选择基期。几何平均数不同于算术平均数，或均值；但它也是合法的计算方法，而且在某些情况下它是一种最有效的方法。计算 3 个数的几何平均数，只需将 3 个数相乘，开 3 次方根；4 个数的几何平均数，开 4 次方根；两个数的则开平方根，以此类推。

以去年为基期为例，也就是说，去年每种商品的价格都看成100％,将两个100％相乘再开平方根,得到100％,这是去年价格指数的几何平均数。今年牛奶是去年的 50％,

面包是去年的 200％，50％乘以 200％得 10000％，再开平方根得 100％。价格没升没降。

事实是，如果不去考虑它的数学基础，统计不仅是一门科学，而且还是一门艺术。许多控制甚至扭曲都是在合理范围之内进行的。统计工作者经常要在许多方法中主观地选择一种方法以描述事实。

在商业活动中，统计工作者不大可能选择不利于己的方法，就像撰稿人在描述赞助商的产品时，不会使用“易碎、价格低贱”的字眼，而会说“轻便、经济、实惠”。

即使是学术界，学者也有自己的偏好(可能是无意识的)。例如，一个需要证明的理论、个人的打算等，都会带有本人的偏好。

所有这些都在提醒大家，在报纸、杂志和书籍中看到统计材料、结论以及数据时，应该经过认真的思考后再接受它们。有时候，更仔细的一瞥有利于进一步了解真相。但武断地拒绝统计方法也是因噎废食，不值得提倡。这类似于仅仅因为作者有时会使用词汇来隐藏而不是揭露事实就拒绝阅读一样。不过，这样的作者也是有的，比如前不久佛罗里达州的某个政客就通过指控他的对手奉行“宗教独身主义”而大赚了一笔。一个在纽约展出电影的人，用极大的字体引用了《时报》对它的评价，称之为“历史的骄傲”。又比如，“神奇的液体水晶”的发明者这样为自己的产品做广告“一种疗效神奇的止痛药”。

第十章、对统计资料提出的五个问题

到目前为止，我一直把自己描述成一个热衷于指导剑术的海盗。在这个总结性章节里，我将摈弃这种文学描述，并直接揭示隐藏在文章表层下的主题：怎样凭双眼就能识破虚假的统计资料并揭穿它；同样重要的是，如何在具有前述欺骗性的数据海洋中找出可靠有用的资料。

你所接触到的统计资料并非都能经受化学分析或者实验室的鉴定。但至少你可以提 5 个简单的问题，在寻找这些问题答案的同时，你将避免接受一些不真实的资料。

谁说的？

首先要寻找的是偏差——出于学说、名誉或收入的考虑而需要证明某些结论的实验室，希望获得一篇好报道的报界，工资已岌岌可危的工人和管理部门都有可能制造偏差。

先来看一看有意识的偏差。这种偏差的表达形式可以是错误的陈述，或许是不易被揭穿的含糊之词。偏差还可以是刻意挑选适合的数据而将不合适的数据放在一边；测量标准的改动也会产生偏差，比如在进行一种比较时确定了某年为基期，而另一种比较却使用了更有利的年份；使用不正确的测量方法也是偏差产生的原因之一，比如说，简单地使用“平均数”一词掩盖了能获得更多信息的中位数被均值替代的事实（或许两者都不能反映实际情况）。

然后，再仔细寻找无意识偏差——通常，它更危险。1928 年，这种偏差使得统计专家和经济学家在制作的图表和进行的预测中得出了惊人的结论。经济结构的不合理被喜悦的情绪忽略了，各种各样的迹象被罗列出来，它们都从统计上支持了这个结论：我们只不过刚刚开始致富。

为了弄清楚究竟是谁做出了结论，即“谁说的”，我们至少应该对资料多看一眼。在很多情况下，《骗术》一书的作者斯蒂芬·珀特(Stephen Potter)所说的“权威人士”掩盖了真实的资料来源。与医药界沾边的任何东西都可以是“权威人士”；科学的实验室也是“权威人士”；大专院校，尤其是大学以及那些在技术方面名列前茅的学校更具有权威性。前几章中那个试图证明高等教育将危及女孩成婚几率的作者，充分地利用了康奈尔大学这个“权威机构”。但请大家注意，虽然数据来自康奈尔大学，结论却完全是作者自己的。康奈尔大学的声望却让你留下了一个错误的印象，“康奈尔大学得出结论……”

当某个权威人士被引用时，请弄清楚到底资料的内容是权威的，还是仅仅与权威人士沾边。

也许你曾经看过芝加哥《商业期刊》(Journal of Commerce)一份很得意的声明。该杂志曾做过一个调查，在 169 家对价格变动的调查做出回答的公司中，有三分之二的公司认为自己正在吞食由朝鲜战争酿成的价格上涨的苦果。“调查显示，”该刊物得出结论(当看到类似的字眼时请格外留意)，“公司的行为与美国商业系统敌人的控诉完全相反。”既然《商业期刊》可能是当事人，我们就应该问一句：“谁说的？”而且我们还应追加第二个问题：

他是如何知道的?

实际上《商业期刊》一开始向 1200 家大公司派发了调查问卷，结果仅仅只有 14％的回收率。也就是说，86％的公司不愿意公开自己在这个问题上的看法。

《商业期刊》过于乐观了，事情并没有什么值得炫耀的。实际情况是：在 1200 家公司中，9％的公司认为价格没有上涨，5％则认为价格上升了，而另外 86％没有回答。以那些做了回答的公司为样本，它可能是有偏的。

注意样本的有偏，是由于选择不当，还是像这个例子一样，由刻意挑选有利的样本造成的。问一个我们前几章曾经提过的问题：

样本是否足够大到能够保证结论值得信赖?

对于相关系数也可以提同样的问题：数值是否足够大从而能说明问题？案例是否足够多？是否具有一定的显著性？作为一个不经意的读者，你也许不会运用显著性检验或根据样本的多少判断结论的准确性。但对于许多报告中的内容，你至少可以进行仔细的观察或许是一个长时间的观察，从而发现由于缺乏足够多的案例，报告的内容不足以说服任何人。

遗漏了什么?

通常，你无法了解样本包含了多少案例。这个数据的缺失，特别当信息的来源存在着利害关系时，已足以使你对整件事情提出质疑。同样，对一个没有经过可信度(可能误差、标准误差)检验的相关也不用太当真。

当均值与中位数相差甚远时，注意那些没有标明类型的平均数。

很多数据因为缺乏比较而变得没有意义。一篇刊登在《嘹望》(look)杂志上关于先天痴呆症的文章指出：“一项研究表明，在 2800 个案例中，超过半数的患者母亲年龄是 35 岁或超过 35 岁。”如果想从中获得信息，读者还应该掌握女性通常的生育年龄，但很少有人了解此类信息。

下面一段文章节选自 1953 年 1 月 31 日《纽约客》(New Yorker)杂志的“伦敦来信”。

卫生部(The Ministry of Health)最近公布的数据表明，在大雾的一周内，伦敦市郊的死亡人数猛增至 2800 人。对于公众来说，这是个极大的震撼。他们常常习惯把英国的天气看成是个麻烦事而不会将之视为杀手……这个冬天大自然的恩赐是不寻常的致命武器……

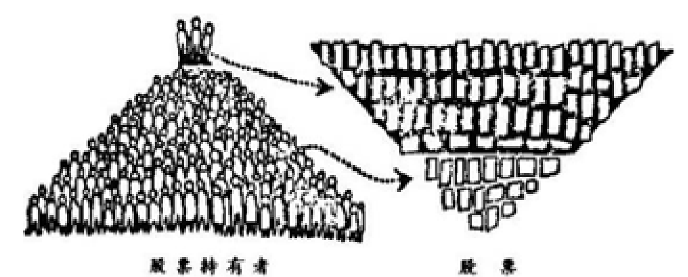
但这个“恩赐”究竟有怎样的杀伤力？我们能否认为该周的死亡率高于平常只是个意外？所有类似的事情都在变动当中，天气在变，死亡人数也在变。接下来几周的情况又如何呢？是否这周的死亡人数低于平均水平就能说明因雾而死的是那些不久于人世的人呢？这个数据是惊人的，但由于缺乏其他数据与之比较，因此就显得毫无意义。

有时仅给出百分数却缺少原始数据也能造成欺骗。很久以前，当约翰斯·霍普金斯大学开始接收女学生时，一个不赞成异性同校的人有一个惊人的报道：约翰斯·霍普金斯大学 1／3 的女学生嫁给了大学老师！然而原始数据更清楚地描绘了事实：总共只有 3 个女同学被录取，其中

1 人嫁给了老师。

几年前，波士顿商会(The Boston Chamber of Commerce)评选出了一些功成名就的妇女，其中 16 位还出现在《美国名人录》中。材料显示她们有“60 个学士学位，18 个孩子”。这已经很清楚了，不是吗？但是当你发现维吉利娅·格尔德丝丽美(Virginia Gildersleeve)校长和丽莲·M.格尔布蕾丝(Lillian M.Gilbreth)夫人也包括在内时，对事实会有更进一步的了解。单单她们两个人就有 20 个学位，而丽莲 ·M.格尔布蕾丝夫人一人就有 12 个孩子。

一家公司可以这样公告，该公司有 3003 名股东，平均每人持股 660 股，的确如此。但下面的说法更清楚，公司总共有 200 万股股票，其中 3 名大股东持有 3／4，而剩下的 3000 人只持有 1／4。



当看到一个指数时，你或许会觉得遗漏了什么。巧妙之处在于基期，一个经过挑选的基期将会扭曲事实。某个全国性的劳工组织曾经证明：经济大萧条之后生产和利润指数上升得比工资指数要快，于是提出增加工资的要求。

但如果人们发现了那些未被披露的数据，就不会提出这样的要求了。利润指数上升得快的原因是因为萧条时期利润指标几乎达到了谷底，于是基期相对较小。

有时文章中遗漏了引起变化的原因，这容易让读者认为其他的因素才应对变化负责。某年公布的数据试图用今年 4 月的零售额高于去年来证明经济处于复苏阶段，而遗漏的内容是去年的复活节在 3 月而今年在 4 月。

某个报告得出了“最近 25 年癌症死亡人数增多”的结论，而这个结论是具有误导性的。让我们看看哪些因素还将影响癌症的死亡人数：以前许多“病因不明”的案例现在已经确诊为癌症；尸体解剖成为一种经常使用的方法，便于做出更确切的诊断；医学统计资料的报告和编制更加全面；易发病年龄段的人数增多。而且，如果你关心死亡总人数更甚于死亡率的话，不要忽略这个事实：现在的人数已远远超过了从前。

是否有人偷换了概念？

分析统计资料时，请留心从搜集原始资料到形成结论的整个过程中，是否存在着概念的偷换。在现实生活中，将甲说成乙的案例实在是太多了。

就像以前曾指出的一样，疾病案例的增多不能等同于发病率的提高，民意测验中的获胜也并不等同于竞选时的获胜，读者对于全球时事文章的偏爱也并不能说明如果杂志刊登此类文章会提高杂志的销售量。

1952 年，加利福尼亚州中部山谷报告的脑炎患者人数是历史最高水平的 3 倍。许多受惊吓的居民将其子女迁出

该州。但如果认真调查，就会发现因脑炎而在熟睡中死去的人数并未增加。导致脑炎“增加”这个假相产生的原因是为了解决该地区一些长期悬而未决的医疗问题，联邦和州的许多医务工作者进入该地区，由于他们的种种努力，许多以前被忽略的或未被发现的低年龄案例都得到了记录。

这不仅让人联想起林肯·斯蒂芬斯(Lincoln Steffen's)和加科博·A.瑞斯(Jacob A.Riis)这两个纽约报界人物几年前制造的犯罪风波。报纸上刊登的犯罪事件一度达到了相当高的水平，报导的犯罪案件不仅数量多，而且占版面的篇幅十分大，加上都用大字标题进行报道，群众纷纷要求采取行动，瑟多·罗斯福(Theodore Roosevelt)——当时的警察联合会主席也陷人了窘境。但他仅仅用一个举动就结束了这场轰轰烈烈的犯罪风波——解雇这两个编辑。事情的起因是他们的手下在相互竞争，看谁能挖掘出更多的犯罪案件，包括抢劫案及诸如此类的案件；而官方的记录显示发案率并没有上升。

“5 岁以上的英国男子在冬天平均每周洗澡 1.7 次，而夏天为 2.1 次。”一篇新闻报道如是说，“英国妇女的相应数据为：冬天 1.5 次，夏天 2.0 次。”该资料来源于英国劳工部(The Ministry of Works)对“6000 户有代表性的英国家庭”所做的调查。看上去样本很具代表性，而且容量也足够大，足以证明旧金山《编年史》(Chronicle)中有趣的结论：“英国的他比她更爱洗澡。”

如果能知道是中位数还是均值，那这个平均数将更有价值。然而，主要的问题还不在于此，而在于概念的偷换。

劳工部所得到的数据只能反映调查者所说的洗澡次数，并不能反映实际洗澡的频率。一个类似洗澡这么私人的问题，并且涉及到英国洗澡的传统时，说的和做的很可能不一致。也许英国的“他”并不比“她”洗澡频繁，一个较安全的结论是他们说的是这样。

偷换概念还有许多值得注意的方式。

最近的一项普查发现，1935 年的农场数目比 5 年前多了 50 多万个，这使得人们认为已经掀起了一场“回归农场”的运动。而实际上，这是由于前后两次的统计口径不同，普查局对农场的定义已经发生了变化，1935 年的口径比 1930 年的范围更广。这样，按照 1930 年口径被排除在外的 30 万个农场被包括进来了。

如果数据是建立在人们口头的回答(即使有一些听上去十分客观)基础之上，将发生许多怪事。比如说，普查报告显示 35 岁的人数远远多于 34 或者 36 岁的人数。这个数据来源于家庭某个成员填报的其他成员的岁数。对于不确定的年龄，人们往往倾向于取 5 的倍数。获得该数据的准确办法是：询问他的出生年月。

某国某个地区的人口总数是 2800 万，5 年后这个数据却变成了 10500 万，当中只有小部分的增长是真实的。产生前后两次巨大差异的原因可归结为不同的调查目的，第一次是为征税和征兵服务的，第二次却是为了发放救济粮。

同样的事在美国也出现过。1950 年的普查显示当时 65～70 岁之间的人数比 10 年前 55～60 岁的人数多。这个增多不是由人口迁移造成的，而主要是因为为了及早享受社会保障金而虚报岁数，或许也有可能是以前为了避免空

虚而少报了岁数。

参议员威廉·兰格(William Langer)的倡议代表了偷换概念的另一种方法。“让一名囚犯离开阿尔卡特拉兹监狱，并将他安置在沃尔多夫—阿斯托里旅馆会更便宜……”这位北达科他州人士在以前的声明中指出，在阿尔卡特拉兹监狱一名囚犯一天的费用是 8 美元，“相当于圣·弗兰西斯科高级酒店的住宿费”。而这里将囚犯的所有生活费与酒店的房租进行比较也是偷换了概念。

在相关事宜中自命不凡的胡说也是一种偷换概念的手法，即将“相关关系”偷换成“因果关系”。《电力世界》(Electrical World)杂志曾经在某篇社论中，公布了一张极其复杂的图表，标题为“电力对美国到底意味着什么？”当“工厂的用电马力”上升时，“平均每小时工资”也在上升，但同时“每周平均工作时间”减少。当然，所有这些都是长期趋势，但并没有证据表明它们当中的某个因素是另一个的起因。

接下来看看“第一”的问题。几乎所有人都可以标榜自己在某个领域获得了第一——如果没有特别指出是哪个领域。1952 年年底两家报纸都坚持自己是百货业广告的老大，从不同的角度来说，两家的结论都是对的。《世界电讯报》(World Telegram)认为在连续刊登广告，即每期都有的广告方面，它是第一，而事实上它也只有这种广告。《美国期刊》(Journal-American)则坚持在刊登整版广告方面它是第一。这种达到最高级的想法会让气象播音员将十分平常的一天说成是“自 1949 年以来，最热的一个 6 月 2 日”。

偷换概念会使你在准备借钱时很难比较直接借款与分期付款的成本。6％听起来就是 6％——但它又有可能是另一回事。

如果你以每月等额分期付款的形式从银行以 6％的利率借了 100 美元，应支付的利息是 3 美元。但如果是另一种 6％，有时也称为借 100 美元还 6 美元，你就需要偿还前者的 2 倍即 6 美元。大多数汽车贷款采用后一种方法，实在是狡猾。

关键在于分期还款时，100 美元你并没有用到一年，6 个月过去后，你已经偿还了一半的贷款。如果采用第二种形式计算利息，即借款额的 6％，你实际承担的利率应该是 12％。更糟糕的是在 1952～1953 年间那些冷冻食物计划的粗心购买者。他们所看到的报价是 6％～12％，看上去这就是利率，但实际却不是，因为它是以美元为基础的，而且更糟的是，期限是半年而不是一年。借 100 美元在每半年后还 12 美元，实际的利率应该是 48％。这就难怪很多消费者不能偿还债务，而且那么多的食品计划泡汤。

人们有时会通过玩文字游戏以达到偷换概念的目的。下面是摘自《商业周刊》(Business Week)杂志的一段话：会计人员往往认为“盈余”是一个会引起诸多争议的词，他们提议把它从公司的资产负债表中抹去。美国会计师协会会计程序委员会建议：……可以使用“留存利润”或者“固定资产增值”这种描述性的术语。

下面这段话摘自一篇报道标准石油公司破记录的收益和一天 100 万净利润的新闻故事：可能董事会会花一点时间来考虑股票分割，因为那会带来一些好处……如果每

股利润看上去不那么大的话……这个资料有意义吗?

当所接触到的资料是建立在未经证实的假设基础之上时，你可以发问，“这个资料有意义吗？”这个问题通常能将一个伪装得很好的统计资料打回原形。你也许十分熟悉鲁道夫·弗列斯基(Rudolf Flesch)的可读性公式。它仅仅依据单词和句子长度这类简单客观的指标就能测量出一篇文章的可读程度。和所有那些将无法估计的事物转化成一个数据或可替代的平均数，从而进行判断的方法一样，它还挺吸引人。至少那些需要雇用作者的人，例如报商，会被它所吸引，即使作者本身对它并不以为然。公式做了这样的假定，诸如单词的长短等因素决定了文章的难易程度，不过，这点还有待证明。

一个名叫罗伯特·A.杜佛(Robert A.Detour)的人曾利用该公式测试了手边一些文学读物。结果证明《睡谷的传说》(The Legend of Sleepy Hollow)的难易程度是柏拉图《共和国》(Republic)的 1.5 倍；辛克莱·刘易斯(Sinclair Lewis)(美国“社会丑事揭发派”作家)的小说《卡斯·廷伯莱恩》(Cass Timmberlane)比雅克·马利丹(Jacques Maritain)的散文《艺术的精神价值》(The Spiritual Value of Art)还要难读。说得跟真的一样。

许多统计资料一眼就可以看出是有误的，这是因为奇妙的数据与平常的感觉不符。莱昂纳多·恩格尔(Leonard Engel)在《琴师》的一篇文章中列举了一些医学上的例子。

其中一个例子是某位著名泌尿学家所做的计算——在美国约有 800 万名前列腺癌症患者——也就是说，易感染的男性群体中，每人有 1.1 个患病机会！另一个例子是

著名的神经专家做出的推测，12 个美国人中有 1 人患有周期性偏头痛，既然周期性偏头痛占了所有头痛病历的三分之一，这就意味着 4 个美国人中有 1 人受到不能思考的头痛困扰。还有一个例子是，复合硬化症患者已达 250000 例，值得庆幸的是，死亡数据却显示全国范围内该病例不超过 30000～40000。

《社会保障法》修正案的听证会常被各种各样的问题所淹没，而这些争论只有当不仔细思考时才会有意义。其中一个争论如下：既然预期寿命大约只有 63 岁，那么将社会保障计划中的退休年龄规定为 65 岁便是惺惺作态。因为所有人在此年龄之前都将死去。

你可以通过观察周围所认识的人进行反驳。问题的关键在于数据指明的是出生时的预期，而出生的婴儿中几乎一半都将活过 63 岁。需要附带说明的是，这个数据来自于最近的官方完整生命表，时间范围是 1939～1941 年。最新的修订数据已改为 65 岁。也许这又会引发另一场同样愚蠢的争论，是否所有人都活到了 65 岁？

几年前一家大型电器设备公司以出生率不断下降为基础(多年以来，大家都是这么认为的)高效地制定出了战后的生产计划，并将小容量家电设备、公寓式的冰箱确定为生产重点。其中一个计划者突然发现了计划与常识的冲突，他用足够长的图表列举了这样的事实：他本人、他的合作伙伴、他的朋友以及他的邻居甚至以前的同学都有了孩子，甚至还打算要 3～4 个孩子。这样规模的家庭不应算作小规模的。这导致了一些开放式的调查与制图，不久这家公司快速地将它的生产重点转移到大家庭使用的电

器上。

让人印象深刻的精确数据也会与现实相左。纽约报纸曾经报道：研究显示，一个与家庭住在一起的上班妇女需要 40.13 美元的日常支出。在读这张报纸时，任何人(只要没有完全丧失逻辑思考能力)就会意识到，物质和精神需要的费用加在一起是不可能精确到美分的。但是，人们又面临着一个致命的诱惑：总认为“40.13 美元”比“大约 40 美元”更精确。

同样，对于几年前美国石油行业委员会(The American Petroleum Industries Committee)的报告：平均每年汽车纳税 51.13 美元时，你也应该表示怀疑。

外推法是十分有用的，特别当预测趋势时。但是，当看到利用外推法计算出来的数据和图表时，请记住这点：到目前为止的趋势都是事实，而未来的趋势只不过是受教育者的猜测。该方法暗含“其他所有条件都相同”，以及“现有趋势将继续下去”的前提。但实际上，条件总是在变化的，否则生活真是无聊透顶了，用一个实例来说明不加控制的外推法内在的不完善性。从 1947～1952 年，美国家庭拥有的电视机数目增加了近 10000％，按照该速率对下一个五年进行推测，你将发现电视机的总数将超过 10 亿台，即每个家庭拥有 40 台电视。更笨的做法是用 1947 年的前两两年为基期，你将“证明”未来家庭拥有的电视机不仅仅是 40 台，而是 4 万台。一个政府研究人员莫瑞斯·汉森(Morris Hansen)，曾将 1948 年嘎勒普的选举预测看成是“人类历史上最具宣传价值的统计错误”。然而，如果将它与广泛使用的人口预测相比，它还真可算得上一

个完美的典范，那个人口预测已经变成了全美国人民的笑柄。1938 年，由专家组成的总统顾问团还在担心美国人口是否可以达到 1.4 亿，结果呢，仅仅过了 12 年，人口总数就超出该数字 1200 万之多。一些近期出版的大学教科书，仍使用了这种预洲并得出结论：未来的美国将不会出现 1.5 亿的人口高峰，即使出现了，也是在 1980 年之后。这些可怕的低估都是以人口趋势保持不变为基础进行预测的。一个世纪以前，由于采用了 1790～1860 年间的人口增长率进行预测，也曾得到过类似的错误结论，只不过这次是高估了人口。在阿伯拉罕·林肯(Abraham Lincoln)递交给国会的第二封信中，预测 1930 年美国人口将达到 251689914 人。

不久后，即 1874 年，马克·吐温(Mark Twain)在《在密西西比河上的生活》(Life on the Mississippi)一书中描述了外推法不严密的特点：在 176 年间，下密西西比河缩短了 242 英里，平均一年 1 又 1/3 英里。因此，任何一个不瞎不傻的正常人都会猜想：明年 11 月份的一百万年以前，即古志留纪时代，下密西西比河有 130 万英里，直接与墨西哥海湾相连，酷似一根钓鱼竿。同理。任何人也将看到再经过 742 年,下密西西比河将变得只有 1 又 3/4 英里长。那时，开罗与新奥尔良的街道将连在一起，人们在同一个市长和相同的市参议员领导下，辛勤而愉快地工作。这就是科学迷人的一面，一个人可以用事实进行微薄的投资却获得如此大的收益！